

## Retrospective Study

# Machine learning-based comparison of factors influencing estimated glomerular filtration rate in Chinese women with or without non-alcoholic fatty liver

I-Chien Chen, Lin-Ju Chou, Shih-Chen Huang, Ta-Wei Chu, Shang-Sen Lee

**Specialty type:** Health care sciences & services**Provenance and peer review:** Unsolicited article; Externally peer reviewed.**Peer-review model:** Single blind**Peer-review report's classification****Scientific Quality:** Grade C, Grade C**Novelty:** Grade B, Grade B**Creativity or Innovation:** Grade B, Grade B**Scientific Significance:** Grade B, Grade C**P-Reviewer:** Wang TJ, China**Received:** January 2, 2024**Revised:** February 13, 2024**Accepted:** April 9, 2024**Published online:** May 26, 2024**I-Chien Chen, Lin-Ju Chou, Shih-Chen Huang,** Department of Nursing, Kaohsiung Armed Forces General Hospital, Kaohsiung 802, Taiwan**Ta-Wei Chu,** Department of Obstetrics and Gynecology, Tri-Service General Hospital, National Defense Medical Center, Taipei 114, Taiwan**Ta-Wei Chu,** Chief Executive Officer's Office, MJ Health Research Foundation, Taipei 114, Taiwan**Shang-Sen Lee,** Department of Urology, Taichung Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, Taichung 427, Taiwan**Shang-Sen Lee,** School of Medicine, Tzu Chi University, Hualien 970, Taiwan**Shang-Sen Lee,** Department of Urology, Tri-Service General Hospital, National Defense Medical Center, Taipei 114, Taiwan**Corresponding author:** Shang-Sen Lee, PhD, Chief Physician, Department of Urology, Taichung Tzu Chi Hospital, Buddhist Tzu Chi Medical Foundation, No. 88 Section 1, Fengxing Road, Tanzi Dist., Taichung 427, Taiwan. [j520037@yahoo.com.tw](mailto:j520037@yahoo.com.tw)

## Abstract

### BACKGROUND

The prevalence of non-alcoholic fatty liver (NAFLD) has increased recently. Subjects with NAFLD are known to have higher chance for renal function impairment. Many past studies used traditional multiple linear regression (MLR) to identify risk factors for decreased estimated glomerular filtration rate (eGFR). However, medical research is increasingly relying on emerging machine learning (Mach-L) methods. The present study enrolled healthy women to identify factors affecting eGFR in subjects with and without NAFLD (NAFLD+, NAFLD-) and to rank their importance.

### AIM

To uses three different Mach-L methods to identify key impact factors for eGFR in healthy women with and without NAFLD.

## METHODS

A total of 65535 healthy female study participants were enrolled from the Taiwan MJ cohort, accounting for 32 independent variables including demographic, biochemistry and lifestyle parameters (independent variables), while eGFR was used as the dependent variable. Aside from MLR, three Mach-L methods were applied, including stochastic gradient boosting, eXtreme gradient boosting and elastic net. Errors of estimation were used to define method accuracy, where smaller degree of error indicated better model performance.

## RESULTS

Income, albumin, eGFR, High density lipoprotein-Cholesterol, phosphorus, forced expiratory volume in one second (FEV1), and sleep time were all lower in the NAFLD+ group, while other factors were all significantly higher except for smoking area. Mach-L had lower estimation errors, thus outperforming MLR. In Model 1, age, uric acid (UA), FEV1, plasma calcium level (Ca), plasma albumin level (Alb) and T-bilirubin were the most important factors in the NAFLD+ group, as opposed to age, UA, FEV1, Alb, lactic dehydrogenase (LDH) and Ca for the NAFLD- group. Given the importance percentage was much higher than the 2<sup>nd</sup> important factor, we built Model 2 by removing age.

## CONCLUSION

The eGFR were lower in the NAFLD+ group compared to the NAFLD- group, with age being was the most important impact factor in both groups of healthy Chinese women, followed by LDH, UA, FEV1 and Alb. However, for the NAFLD- group, TSH and SBP were the 5<sup>th</sup> and 6<sup>th</sup> most important factors, as opposed to Ca and BF in the NAFLD+ group.

**Key Words:** Non-alcoholic fatty liver; Estimated glomerular filtration rate; Machine learning; Chinese women

©The Author(s) 2024. Published by Baishideng Publishing Group Inc. All rights reserved.

**Core Tip:** We examined influential factors affecting the estimated glomerular filtration rate in healthy women with and without non-alcoholic fatty liver disease (NAFLD) by multiple linear regression and machine learning methods, with machine learning methods providing better performance and showing that age was the most important determining factor in both groups, followed by lactic dehydrogenase, uric acid, forced expiratory volume in one second, and albumin. However, for the NAFLD- group, the 5<sup>th</sup> and 6<sup>th</sup> most important impact factors were thyroid-stimulating hormone and systolic blood pressure, as compared to plasma calcium and body fat for the NAFLD+ group.

**Citation:** Chen IC, Chou LJ, Huang SC, Chu TW, Lee SS. Machine learning-based comparison of factors influencing estimated glomerular filtration rate in Chinese women with or without non-alcoholic fatty liver. *World J Clin Cases* 2024; 12(15): 2506-2521

**URL:** <https://www.wjgnet.com/2307-8960/full/v12/i15/2506.htm>

**DOI:** <https://dx.doi.org/10.12998/wjcc.v12.i15.2506>

## INTRODUCTION

The increased Westernization of lifestyles in Taiwan has driven the prevalence of obesity and obesity-related diseases such as diabetes and metabolic syndrome[1]. Total caloric intake exceeding total energy expenditure results in the accumulation of fat in the liver, leading to nonalcoholic fatty liver disease (NAFLD)[2]. The prevalence of NAFLD has increased recently, ranging from 8%-45% in various countries[3]. Around one-fourth of NAFLD cases eventually progresses to nonalcoholic steatohepatitis, with another 20%-35% further progressing to fibrosis and even hepatocellular carcinoma[4]. Thus, NAFLD is a critical issue for both the community and health providers.

NAFLD has also been shown to be associated with chronic renal disease (CKD)[5], with insulin resistance, obesity, dyslipidemia, and hypertension found to contribute to impaired renal function[5-8]. However, most of these studies relied on traditional statistical analysis methods, with machine learning approaches use in only one study to identify four NAFLD-related genes[9].

Medical research is increasingly applying machine learning (Mach-L) methods to improve prediction accuracy by repeated analysis of large data sets. Some of these studies focused in this area began to use Mach-L[10]. The present study uses three different Mach-L methods to identify key impact factors for estimated glomerular filtration rate (eGFR) in healthy women with and without NAFLD.

## MATERIALS AND METHODS

### Participant and study design

Data for this study were sourced from the Taiwan MJ Cohort, an ongoing prospective cohort of health examinations conducted by the MJ Health Screening Centers (MJ) a group of three private clinics in Taiwan that provide regular health examinations to their members[11]. These examinations cover more than 100 important biological indicators, including anthropometric measurements, blood tests, imaging tests, *etc.* Each participant completed a self-administered questionnaire covering personal and family medical history, current health status, lifestyle, physical exercise, sleep habits, and dietary habits[12]. The MJ Health Database only includes participants who have provided informed consent. All or part of the data used in this research were authorized by and received from the MJ Health Research Foundation (Authorization Code: MJHRF2023008A). Any interpretations or conclusions described in this paper are those of the authors alone, and do not represent the views of the MJ Health Research Foundation. The study protocol was approved by the Institutional Review Board of the Kaohsiung Armed Forces General Hospital (IRB No.: KAFGHIRB 112-007). As a secondary study which does not collect samples from study participants, a short review was applied and no further informed consent was needed.

A total of 102 847 healthy women were initially enrolled. After excluding subjects with different causes, 65535 women remained for analysis, as shown in **Figure 1**. The inclusion criteria were as follows: (1) Age between 18 to 65 years old; (2) No known significant disease such as coronary heart disease, stroke, or chronic renal disease; (3) Not taking medications for hypertension, hyperlipidemia, or hyperglycemia; and (4) No regular alcohol consumption; In Model 1, the participants were divided into groups with or without NAFLD (NAFLD+, NAFLD-, respectively). Preliminary Mach-L analysis found that age was the most important factor in both groups (mean = 71.69%), followed by uric acid (respectively 37.17% and 41.99% in the NAFLD+ and NAFLD- groups). We made the second models (Model 2) by removing age from the data set.

The following methods were published in our previous study[13]. On the day of the study, senior nursing staff recorded the subject's medical history, including information on any current medications, and a physical examination was performed. The waist circumference was measured horizontally at the level of the natural waist. The body mass index (BMI) was calculated as the participant's body weight (kg) divided by the square of the participant's height (m). Systolic blood pressure (SBP) and diastolic blood pressure were measured using standard mercury sphygmomanometers on the right arm of each subject while seated.

After fasting for 10 h, blood samples were collected for biochemical analyses. Plasma was separated from the blood within 1 h of collection and stored at 30 °C until analysis for fasting plasma glucose (FPG) and lipid profiling. FPG was measured using the glucose oxidase method (YSI 203 glucose analyzer; Yellow Springs Instruments, Yellow Springs, OH, United States). The triglyceride (TG) levels were measured using the dry multilayer analytical slide method with a Fuji Dri-Chem 3000 analyzer (Fuji Film, Tokyo, Japan). The serum high-density lipoprotein cholesterol (HDL-C) and low-density lipoprotein cholesterol (LDL-C) concentrations were analyzed using an enzymatic cholesterol assay, following dextran sulfate precipitation. The BMD was measured by dual-energy X-ray absorptiometry (Lunar, General Electric, United States). Fatty liver was diagnosed based on the following ultrasound parameters: Parenchymal brightness, liver-to-kidney contrast, deep beam attenuation, bright vessel walls, and gallbladder wall definition. Qualitative grades are conveniently labeled mild (3), moderate (2), severe (1) or normal (0). In the present study, grades 1 to 3 were all defined as fatty liver (FA).

**Table 1** shows the 32 variables which were used in the present study. These included participants' body fat (BF), complete blood cell count, biochemistries, thyroid stimulating hormone, C-reactive protein (CRP), education level, marital status, and income level. Drinking area was defined as the product of total drinking duration, frequency of drinking and alcohol percentage. Similarly, smoking area was the product of the duration, frequency of smoking and number of cigarettes. The sport area was the product of duration, frequency, and type of exercise. All the aforementioned parameters were the independent variables and eGFR was the dependent variable.

### Traditional statistical method

The data were tested for normal distribution by using the Kolmogorov-Smirnov test and for the homogeneity of variances using Levene's test. Continuous variables were expressed as mean  $\pm$  SD. One way analysis of variance was applied to compare differences between groups (> 2 groups). The Bonferroni test was for post-hoc evaluation. Pearson's correlation was used to examine the relationships between age and other continuous variables.

### Machine learning method

Predictive models for NAFLD with factor ranking were constructed using three different Mach-L methods and part of the following information was published by our group previously[13].

Stochastic gradient boosting (SGB) is a tree-based gradient boosting learning algorithm that combines both bagging and boosting techniques to minimize the loss function to solve the overfitting problem of traditional decision trees[14]. In SGB, many stochastic weak learners of trees are sequentially generated through multiple iterations, where each tree concentrates on correcting or explaining errors in the tree generated in the previous iteration. That is, the residual of the previous iteration tree is used as the input for the newly generated tree. This iterative process is repeated until the convergence condition or a stopping criterion is reached for the maximum number of iterations. Finally, the cumulative results of many trees are used to determine the final robust model.

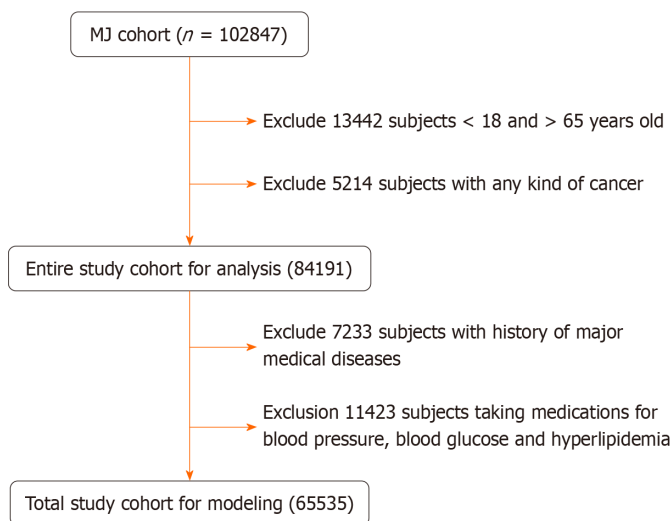
The second method used in this study is eXtreme gradient boosting (XGBoost), which is based on an optimized extension of SGB[15]. It sequentially trains multiple weak models and assembles them using the gradient boosting output

**Table 1 Participant descriptive statistics and risk factors, percentage and mean ( $\pm$  SD)**

Characteristics	None	Fatty liver	P value
N number	34, 335	31, 200	
Age (yr)	36.75 $\pm$ 12.33	47.52 $\pm$ 12.8	< 0.001
Income	2.04 $\pm$ 1.47	1.61 $\pm$ 1.57	< 0.001
Body fat (%)	26.66 $\pm$ 5.55	35.96 $\pm$ 6.83	< 0.001
Systolic blood pressure (mmHg)	111.83 $\pm$ 16.07	124.52 $\pm$ 19.68	< 0.001
Diastolic blood pressure (mmHg)	66.81 $\pm$ 10.21	73.8 $\pm$ 11.69	< 0.001
Leukocyte ( $\times 10^3/\mu\text{L}$ )	5.93 $\pm$ 1.73	6.45 $\pm$ 1.73	< 0.001
Hemoglobin ( $\times 10^6/\mu\text{L}$ )	13.09 $\pm$ 1.14	13.36 $\pm$ 1.18	< 0.001
Platelets ( $\times 10^3/\mu\text{L}$ )	248.96 $\pm$ 57.74	264.39 $\pm$ 63.31	< 0.001
Fasting plasma glucose (mg/dL)	92.75 $\pm$ 10.64	103.7 $\pm$ 25.6	< 0.001
Total bilirubin (mg/dL)	0.77 $\pm$ 0.32	0.73 $\pm$ 0.33	< 0.001
Albumin (mg/dL)	4.5 $\pm$ 0.26	4.45 $\pm$ 0.24	< 0.001
Globulin (mg/dL)	3.08 $\pm$ 0.36	3.15 $\pm$ 0.36	< 0.001
Alkaline Phosphatase (IU/L)	101.86 $\pm$ 49.29	110.72 $\pm$ 59.94	< 0.001
Serum glutamic oxaloacetic transaminase (mg/dL)	19.88 $\pm$ 11.71	24.61 $\pm$ 17.56	< 0.001
Serum glutamic pyruvic transaminase (IU/L)	17.68 $\pm$ 18.62	28.72 $\pm$ 27.22	< 0.001
Serum $\gamma$ -glutamyl transpeptidase (IU/L)	14.22 $\pm$ 13.94	25.3 $\pm$ 30.09	< 0.001
Lactate dehydrogenase (IU/L)	241.42 $\pm$ 84.03	246.65 $\pm$ 92.63	< 0.001
Estimated glomerular filtration rate (mL/min/1.73 m <sup>2</sup> )	89.6 $\pm$ 76.89	84.33 $\pm$ 72.4	< 0.001
Uric acid (mg/dL)	4.88 $\pm$ 1.09	5.67 $\pm$ 1.34	< 0.001
Triglyceride (mg/dL)	78.81 $\pm$ 42.95	139.02 $\pm$ 101.03	< 0.001
High density lipoprotein cholesterol (mg/dL)	61.73 $\pm$ 14.8	54.17 $\pm$ 13.47	< 0.001
Low density lipoprotein cholesterol (mg/dL)	107.97 $\pm$ 30.48	125.76 $\pm$ 34.09	< 0.001
Calcium (mg/dL)	9.2 $\pm$ 0.39	9.29 $\pm$ 0.41	< 0.001
Phosphorus (mg/dL)	3.72 $\pm$ 0.44	3.7 $\pm$ 0.46	< 0.001
Thyroid stimulating hormone (IU/mL)	1.75 $\pm$ 3.25	1.94 $\pm$ 3.64	< 0.001
C-reactive protein (mg/dL)	0.19 $\pm$ 0.45	0.3 $\pm$ 0.51	< 0.001
Forced expiratory volume in one second (L)	2.2 $\pm$ 0.46	1.96 $\pm$ 0.53	< 0.001
Drink area	0.97 $\pm$ 6.5	1.38 $\pm$ 8.72	< 0.001
Smoke area	1.5 $\pm$ 7.15	1.57 $\pm$ 7.89	< 0.001
Betel nut area	0 $\pm$ 0	0.02 $\pm$ 1.2	< 0.001
Sport area	3.32 $\pm$ 5.99	3.96 $\pm$ 6.19	< 0.001
Sleep time	2.91 $\pm$ 0.59	2.87 $\pm$ 0.72	0.25
Marriage, <i>n</i> (%)			
Unmarried	13 (458)	8 (438)	< 0.001
Married	19 (939)	2 (1545)	

method, thus improving prediction performance. XGBoost uses Taylor's binomial expansion to approximate the objective function and arbitrary differentiable loss functions to accelerate the model construction convergence process[16]. XGBoost also applies regularized boosting techniques to penalize model complexity and correct overfitting, thereby increasing model's accuracy[15].

Finally, elastic net (EN) can be regarded as a hybrid version of L1 regularization and L2 regularization. The elastic network uses both L1 and L2 regularization, thus integrating the penalty terms of L1 and L2. The advantage of the EN model is that it combines the Ridge penalty item to achieve effective regularization and the Lasso penalty item to select



**Figure 1** Flowchart of sample selection from the MJ Health Screening Centers Study Cohort.

variables, allowing it to learn models with only a small number of arguments that are non-zero sparse, just like Lasso, but it still maintains some of Ridge's regular properties. EN provides several advantages as follows: It encourages group effects in the case of highly correlated variables, rather than setting some of them to 0 Like Lasso; EN is useful when multiple features are mutually correlated; and Lasso tends to choose one of them at random, while EN tends to choose two[17].

Figure 2 presents the proposed prediction and important variable identification scheme that combines the three Mach-L methods. First, patient data were collected to prepare the dataset. The dataset was then randomly divided into an 80% training dataset for model building and a 20% testing dataset for model testing. In the training process, the hyperparameters of each Mach-L method must be tuned to ensure model effectiveness. In this study, a 10-fold cross-validation technique was used for hyperparameter tuning.

To this end, the training dataset was further randomly divided into a training dataset to build the model with different sets of hyperparameters, and a validation dataset for model validation. All possible combinations of hyperparameters were investigated by grid search. The model with the lowest root mean square error on the validation dataset was viewed as the best model of each Mach-L method. The best models for SGB, XGBoost and EN were generated, and the corresponding variable importance ranking information obtained.

During the testing phase, the performance of the best machine learning models was evaluated using the testing dataset. Since the target variable in this study is a numerical variable, model performance was compared using different metrics, including mean absolute percentage error (MAPE), symmetric mean absolute percentage error (SMAPE), relative absolute error (RAE), root relative squared error (RRSE), and root mean squared error (RMSE). The values for these metrics are listed in Figure 3.

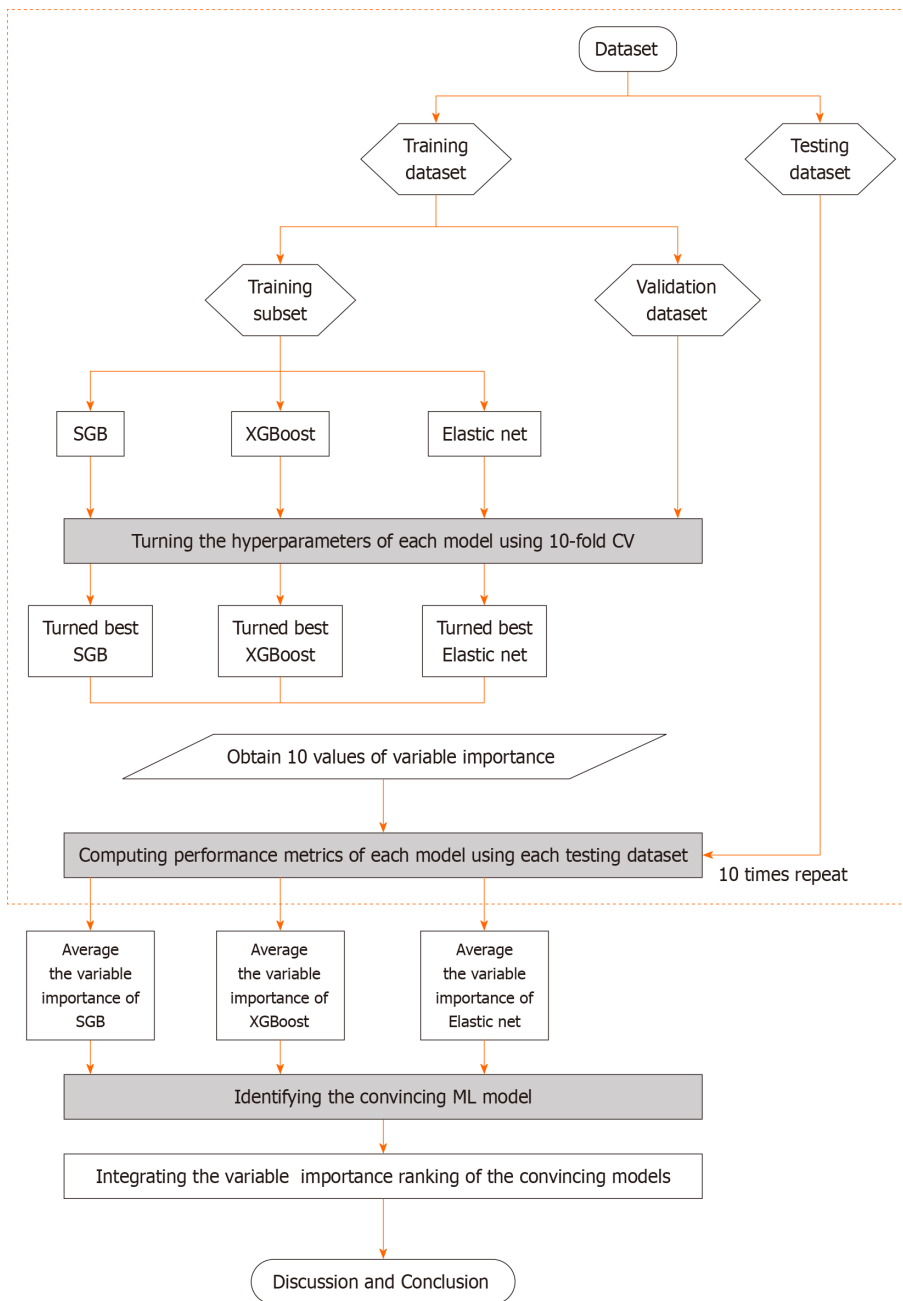
To maximize comparison reliability and stability, the training and testing processes were repeated 10 times. The performance metrics of the three machine learning models were then averaged to compare with the performance of the benchmark multiple linear regression (MLR) model. The same training and testing datasets were used for three the machine learning methods and the MLR model. A model with an average metric lower than that of the MLR model was considered a more convincing model.

The Mach-L methods used in this study may produce percentage of importance of variables due to their unique modeling characteristics. In the final stage of our proposed scheme, we summarize and discuss our significant findings based on the results of Mach-L methods.

All methods were performed using R software version 4.0.5 and RStudio version 1.1.453 with the required packages installed (<http://www.R-project.org>, accessed on; <https://www.rstudio.com/products/rstudio/>).

## RESULTS

The present study enrolled a total of 65535 participants. Table 1 compares all the variables in the NAFLD+ and NAFLD- groups. In the NAFLD+ group, the following variables were all significantly higher: Age, BF, blood pressure, leukocyte, hemoglobin, platelets, FPG, globulin, alkaline phosphatase, serum glutamic oxaloacetic transaminase, serum glutamic pyruvic transaminase, serum  $\gamma$ -glutamyl transpeptidase, lactic dehydrogenase (LDH), uric acid (UA), Total bilirubin (T-bili), TG, LDL-C, Calcium (Ca), Thyroid stimulating hormone (TSH), CRP, drinking area, betel nut area, and sport area. In the NAFLD+ group, the following variables were significantly lower: Income, albumin (Alb), eGFR, HDL-C, Phosphorus, forced expiratory volume in one second (FEV1), and sleep time. The  $\chi^2$  test results showed married participants were more likely to have NAFLD.



**Figure 2** Flowchart of the proposed machine learning methods. XGBoost: eXtreme gradient boosting.

Table 2 compares the performance of LR and other Mach-L methods. Both in Model 1 (with age) and Model 2 (without age), most of the SMAPE, RAE, RRSE and RMSE values were lower in the three Mach-L methods, indicating that the Mach-L methods outperformed MLR. Tables 3-6 show the he importance percentages of the three Mach-L methods, with the averages of these three methods presented in the right column, showing that the most important factors predicting eGFR were age, UA FEV1, Ca, Alb and T-bili in the NAFLD+ group and age, UA, FEV1, Alb, LDH and Ca in the NAFLD-group. The importance percentage of age (71.73% and 71.69%, respectively, for the NAFLD- and NAFLD+ groups) is much higher than that of the following factor, UA (41.99% and 37.17%, respectively for the NAFLD- and NAFLD+ groups). Model 2 was built by removing the age variable, with results shown in Tables 5 and 6. Similar to Model 1, the most important factors were LDH, UA, FEV1, Alb TSH, and SBP for NAFLD- and LDH, UA, FEV1, Alb, Ca and BF for NAFLD+. In Tables 3-6, it should be noted that we used grayscale to show the importance of these factors, the darker the shade of gray, the more important that particular variable was. Figure 4 are the graphic illustrations for Tables 3-6. In Figure 4A, the percentage of importance percentage of the risk factors in participants with NAFLD are shown (Model 1: with age). In the same time, Figure 4B shows the results of without NAFLD in Model 1. Respectively, Figure 4C are results of Model 2 with NAFLD and 3-4 without NAFLD. All the orders of the ranks of percentage importance were the same as the results in the tables. These figures could give a direct concept of these risk factors.

**Table 2 Comparison with SMAPE, RAE, RRSE, and RMSE between multiple linear regression and machine learning methods**

NAFLD+ group with age	MAPE	SMAPE	RAE	RRSE	RMSE
Linear	0.139	0.132	0.845	0.842	13.959
SGB	0.138	0.131	0.841	0.834	13.825
XGBoost	0.139	0.132	0.845	0.842	13.946
Elasticnet	0.139	0.132	0.845	0.842	13.954
<b>NAFLD- group with age</b>					
Linear	0.133	0.128	0.868	0.862	14.671
SGB	0.132	0.126	0.855	0.857	14.59
XGboost	0.132	0.126	0.853	0.857	14.58
Elasticnet	0.134	0.128	0.868	0.862	14.673
<b>NAFLD+ group without age</b>					
Linear	0.154	0.14	0.872	0.897	15.606
SGB	0.153	0.139	0.865	0.888	15.444
XGboost	0.153	0.14	0.869	0.891	15.49
Elasticnet	0.154	0.14	0.872	0.897	15.596
<b>NAFLD- group without age</b>					
Linear	0.134	0.13	0.905	0.906	15.149
SGB	0.133	0.129	0.895	0.892	14.915
XGboost	0.133	0.129	0.895	0.893	14.916
Elasticnet	0.134	0.13	0.904	0.905	15.119

Data showed as mean; SGB: stochastic gradient boosting; XGBoost: eXtreme gradient boosting; RAE: Relative Absolute Error; RRSE: Root Relative Squared Error; RMSE: Root Mean Squared Error; NAFLD: Non-alcoholic fatty liver.

Metrics	Description	Calculation
SMAPE	Symmetric Mean Absolute Percentage Error	$SMAPE = \frac{1}{n} \sum_{i=1}^n \frac{ y_i - \hat{y}_i }{( y_i  +  \hat{y}_i )/2} \times 100$
MAPE	Mean Absolute Percentage Error	$MAPE = \frac{1}{n} \sum_{i=1}^n \left  \frac{y_i - \hat{y}_i}{y_i} \right  \times 100$
RAE	Relative Absolute Error	$RAE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i)^2}}$
RRSE	Root Relative Squared Error	$RRSE = \sqrt{\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \hat{y}_i)^2}}$
RMSE	Root Mean Squared Error	$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$

**Figure 3 Equation of performance evaluation metrics.**  $\hat{y}_i$  and  $y_i$  respectively represent predicted and actual values;  $n$  is the number of instances. SMAPE: Symmetric Mean Absolute Percentage Error; MAPE: Mean Absolute Percentage Error; RAE: Relative Absolute Error; RRSE: Root Relative Squared Error; RMSE: Root Mean Squared Error.

## DISCUSSION

We first demonstrated that the eGFR was lower in NAFLD+ group, and identified the six most important factors for NAFLD- and NAFLD+ in Model 1. Model 2 was then built by removing age. Consistent with Model 1, in Model 2 the first four factors were LDH, UA, FEV1 and Alb, but the 5<sup>th</sup> and 6<sup>th</sup> most important factors were respectively TSH and SBP in NAFLD- and Alb, BF in NAFLD+.

**Table 3** The average of the importance of risk factors derived from stochastic gradient boosting, random forest and extreme gradient boost, in NAFLD+ (Model 1, including age)

Variables	SGB	XGBoost	Elasticnet	Average	Rank
Age	100	100	15.08	71.69	1
Income	0.14	0	0	0.05	
Body fat	3.94	1.9	3.27	3.04	
Systolic blood pressure	1.37	0.67	1.01	1.02	
Diastolic blood pressure	2.67	0.67	1.8	1.71	
Leukocyte	0.72	0.33	4.32	1.79	
Hemoglobin	1	0	0	0.33	
Platelets	3.64	1.62	0.31	1.86	
Fasting plasma glucose	2.92	1.18	0.66	1.59	
Total bilirubin	6.86	2.29	40.24	16.46	6
Albumin	1.84	0.54	49.85	17.41	5
Globulin	0.28	0.29	17.83	6.13	
Alkaline Phosphatase	0.99	0.15	0	0.38	
Serum glutamic oxaloacetic transaminase	1.63	0.36	0	0.66	
Serum glutamic pyruvic transaminase	3.83	1.94	0.82	2.20	
Serum $\gamma$ -glutamyl transpeptidase	1.89	1.33	0.48	1.23	
Lactate dehydrogenase	23.21	23.25	0.86	15.77	
Uric acid	27.03	24.05	60.42	37.17	2
Triglyceride	0.84	0	0.02	0.29	
High density lipoprotein cholesterol	1.99	0.8	1.16	1.32	
Low density lipoprotein cholesterol	1.78	0.14	0.1	0.67	
Calcium	3.97	2.87	65	23.95	4
Phosphorus	0.79	0.36	0	0.38	
Thyroid stimulating hormone	6.92	3.99	4.91	5.27	
C-reactive protein	0.61	0.22	6.99	2.61	
Forced expiratory volume in one second	6.63	3.41	100	36.68	3
Drink area	0.11	0	0.07	0.06	
Smoke area	0.25	0	0	0.08	
Betel nut area	0	0	0	0.00	
Sport area	0.45	0.2	0.28	0.31	
Sleep time	0.14	0	0	0.05	
Marriage	0	0	1.99	0.66	

SGB: Stochastic gradient boosting; XGBoost: eXtreme gradient boosting; NAFLD: Non-alcoholic fatty liver.

Age was found to be the most important impact factor for eGFR. This is not surprising since renal function progress decreases with aging due to the reduction in the glomerular capillary plasma flow rate and its ultrafiltration coefficient. At the same time, decreased afferent arteriolar resistance directly increases capillary hydraulic pressure. These derangements will induce loss of renal mass, hyalinization, sclerotic glomeruli, and tubulointerstitial fibrosis[18].

Interestingly, Model 2 shows that the order of first four factors (LDH, UA, FEV1 and Alb) were identical in participants with or without NAFLD. The implications of this are detailed as follows.

LDH: LDH is an enzyme in cells which converts glucose to pyruvate under aerobic conditions and is then oxidized to acetyl-CoA. It enters the tricarboxylic acid cycle in mitochondria to generate energy. The lactate signaling axis (LDH-lactate-lactation) has been shown to have many physiological roles related to diseases[19,20]. LDH is determined in



**Table 4** The average of the importance of risk factors derived from stochastic gradient boosting, random forest and extreme gradient boost, in NAFLD- (Model 1, including age)

Variables	SGB	XGBoost	Elasticnet	Average	Rank
Age	100	100	15.19	71.73	1
Income	0	0	1.41	0.47	
Body fat	3.69	1.05	4.15	2.96	
Systolic blood pressure	0.46	0.09	0.75	0.43	
Diastolic blood pressure	4.19	3.09	2.96	3.41	
Leukocyte	1.21	0.34	4.52	2.02	
Hemoglobin	4.81	1	11.57	5.79	
Platelets	2.61	1.06	0.2	1.29	
Fasting plasma glucose	0.42	0.21	0.17	0.27	
Total bilirubin	3.11	1.84	19.24	8.06	
Albumin	2.28	1.34	69.53	24.38	4
Globulin	0.42	0.12	3.03	1.19	
Alkaline Phosphatase	1.84	0.22	0.04	0.70	
Serum glutamic oxaloacetic transaminase	0.52	0	0	0.17	
Serum glutamic pyruvic transaminase	3.79	1.94	1.12	2.28	
Serum $\gamma$ -glutamyl transpeptidase	1.16	0	0.38	0.51	
Lactate dehydrogenase	21.99	18.24	0.97	13.73	5
Uric acid	26.62	22.99	76.35	41.99	2
Triglyceride	1.59	0.31	0	0.63	
High density lipoprotein cholesterol	1.4	0.25	0.65	0.77	
Low density lipoprotein cholesterol	2.37	0.26	0.18	0.94	
Calcium	1.66	0.64	29.96	10.75	6
Phosphorus	2.05	2.07	10.98	5.03	
Thyroid stimulating hormone	11.86	8.69	5.02	8.52	
C-reactive protein	0.42	0	2.73	1.05	
Forced expiratory volume in one second	5.06	3.01	100	36.02	3
Drink area	0	0.25	0	0.08	
Smoke area	0.37	0.17	0.79	0.44	
Betel nut area	0	0	0	0.00	
Sport area	1.13	0.65	1.55	1.11	
Sleep time	0	0	0	0.00	
Marriage	0.13	0	5.51	1.88	

SGB: Stochastic gradient boosting; XGBoost: eXtreme gradient boosting; NAFLD: Non-alcoholic fatty liver.

different zones of normal kidneys and patients with impaired renal function. In 1968, Nielsen *et al*[21] reported higher LDH levels in patients with chronic renal failure, due to by changes of LDH synthesis in response to different types of kidney injury. The results of the present study indicate that, after removing age, LDH was the most important factor related to eGFR.

UA: UA is an end-product of purine metabolism in humans and is mostly excreted *via* the kidneys. Elevated UA is related to many diseases such as gout, diabetes, and metabolic syndrome[22-24]. It could also impair renal function and eventually lead to chronic renal disease *via* various mechanisms such as endothelial dysfunction, activation of renin-angiotensin system, inflammation, and oxidation stress[25-29]. In the present study, it is the second-most important impact factor in Model 2.

**Table 5** The average of the importance of risk factors derived from stochastic gradient boosting, random forest and extreme gradient boost, in NAFLD+ (Model 2, excluding age)

Variables	SGB	XGBoost	Elasticnet	Average	Rank
Income	12.88	15.99	8.06	12.31	
Body fat	22.25	16.83	3.85	14.31	5
Systolic blood pressure	11.29	9.24	0.41	6.98	
Diastolic blood pressure	9.17	6.85	1.19	5.74	
Leukocyte	1.73	0.58	1.9	1.40	
Hemoglobin	2.57	0.31	4.27	2.38	
Platelets	17.83	14.87	0.32	11.01	
Fasting plasma glucose	9.72	6.71	0.2	5.54	
Total bilirubin	9.6	3.56	28.65	13.94	
Albumin	12.05	9.81	100	40.62	3
Globulin	1.85	1.25	10.62	4.57	
Alkaline Phosphatase	4.28	0.99	0.06	1.78	
Serum glutamic oxaloacetic transaminase	3.45	3.05	1.45	2.65	
Serum glutamic pyruvic transaminase	16.27	11.92	1.57	9.92	
Serum $\gamma$ -glutamyl transpeptidase	1.14	0.65	0.28	0.69	
Lactate dehydrogenase	100	100	0.6	66.87	1
Uric acid	50.16	45.66	30.15	41.99	2
Triglyceride	6.23	3.56	0.14	3.31	
High density lipoprotein cholesterol	0.86	0.82	0.05	0.58	
Low density lipoprotein cholesterol	9.6	6.73	0.46	5.60	
Calcium	12.48	9.07	53.48	25.01	4
Phosphorus	0.79	1.87	1.48	1.38	
Thyroid stimulating hormone	16.42	11.23	2.23	9.96	
C-reactive protein	0	0	2.11	0.70	
Forced expiratory volume in one second	39.39	44.32	38.15	40.62	3
Drink area	0	0	0.26	0.09	
Smoke area	0	0	0	0.00	
Betel nut area	0	0	0	0.00	
Sport area	3.95	3.83	1.49	3.09	
Sleep time	0.86	0.5	4.04	1.80	
Marriage	0	0	2.38	0.79	

SGB: Stochastic gradient boosting; XGBoost: eXtreme gradient boosting; NAFLD: Non-alcoholic fatty liver.

FEV1: FEV1 is one of the key parameters for evaluating pulmonary function[30] and is the third-most important impact factor in the present study. The relationship between pulmonary function and renal function is well-established. For example, by using data from National Health and Nutrition Examination Survey between 2007 to 2012, Navaneethan *et al* showed that an estimated glomerular filtration rate (eGFR)  $< 60$  mL/min/1.73 m<sup>2</sup> was associated with higher odds of obstructive lung function[31]. Yu *et al*[32] found that in both Australian and Chinese populations, subjects with lower FEV1 Levels generally had normal renal function. Our results are consistent with these findings.

Alb: Alb is one kind of protein in humans. It consists about half of total protein in the serum, and is used as a biomarker for many diseases. Decreased Alb levels are associated with cardiovascular disease, hear failure, and higher mortality rate[33-35]. Similar to pulmonary function, many studies have found that Alb levels are associated with renal function. For example, Lang *et al*[36] found that lower Alb levels were independently and significantly related to impaired renal function (0.11 mL/min/1.73 m<sup>2</sup> per year for each standard deviation fall in Alb) in the elderly Americans,

**Table 6** The average of the importance of risk factors derived from stochastic gradient boosting, random forest and extreme gradient boost, in NAFLD- (Model 2, excluding age)

Variables	SGB	XGBoost	Elasticnet	Average	Rank
Income	2.49	1.95	1.61	2.02	
Body fat	7.57	2.68	1.48	3.91	
Systolic blood pressure	28.66	30.68	0.59	19.98	6
Diastolic blood pressure	18.44	21.96	1.73	14.04	
Leukocyte	9.07	5.26	6.63	6.99	
Hemoglobin	12.51	1.95	3.14	5.87	
Platelets	12.13	8.68	0.23	7.01	
Fasting plasma glucose	6.67	4.96	0.7	4.11	
Total bilirubin	9.07	5.16	3.37	5.87	
Albumin	21.95	20.16	100	47.37	4
Globulin	1.32	0	0	0.44	
Alkaline Phosphatase	2.75	0	0.02	0.92	
Serum glutamic oxaloacetic transaminase	4.06	3.15	1.59	2.93	
Serum glutamic pyruvic transaminase	9.09	6.48	1.74	5.77	
Serum $\gamma$ -glutamyl transpeptidase	1.11	0	0.11	0.41	
Lactate dehydrogenase	100	100	0.63	66.88	1
Uric acid	66.92	63.24	36.68	55.61	2
Triglyceride	12.39	8	0.34	6.91	
High density lipoprotein cholesterol	2.64	0.67	0.17	1.16	
Low density lipoprotein cholesterol	14.18	10.11	0.51	8.27	
Calcium	3.82	2.4	21.04	9.09	
Phosphorus	5.65	6.52	0.1	4.09	
Thyroid stimulating hormone	34.32	24.16	2.56	20.35	5
C-reactive protein	2.68	0	0	0.89	
Forced expiratory volume in one second	61.02	64.4	26.93	50.78	3
Drink area	1.01	1.21	0	0.74	
Smoke area	2.24	1.22	0.85	1.44	
Betel nut area	0	0	0	0.00	
Sport area	13.32	11.44	2.46	9.07	
Sleep time	0.79	0.47	10.69	3.98	
Marriage	8.88	7.85	30.55	15.76	

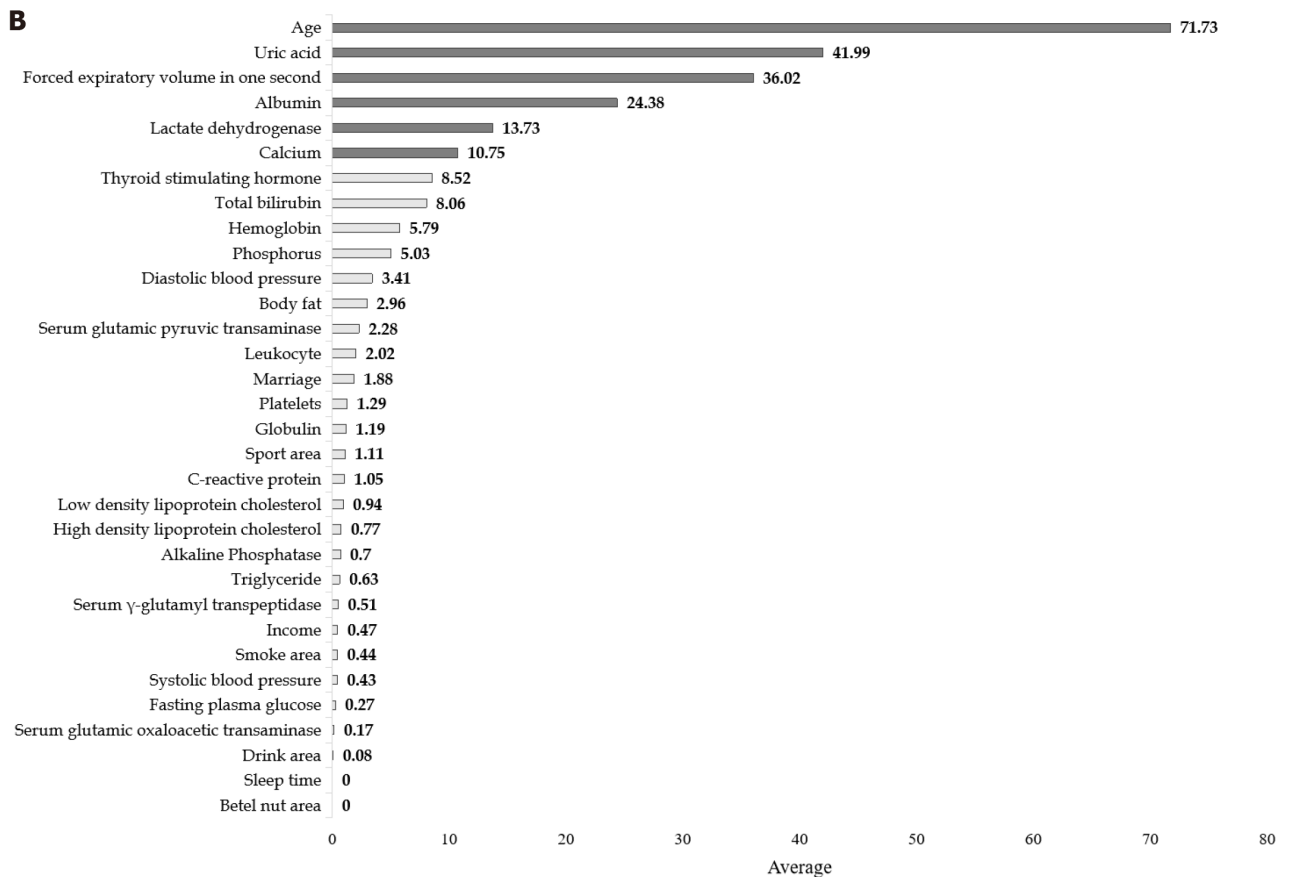
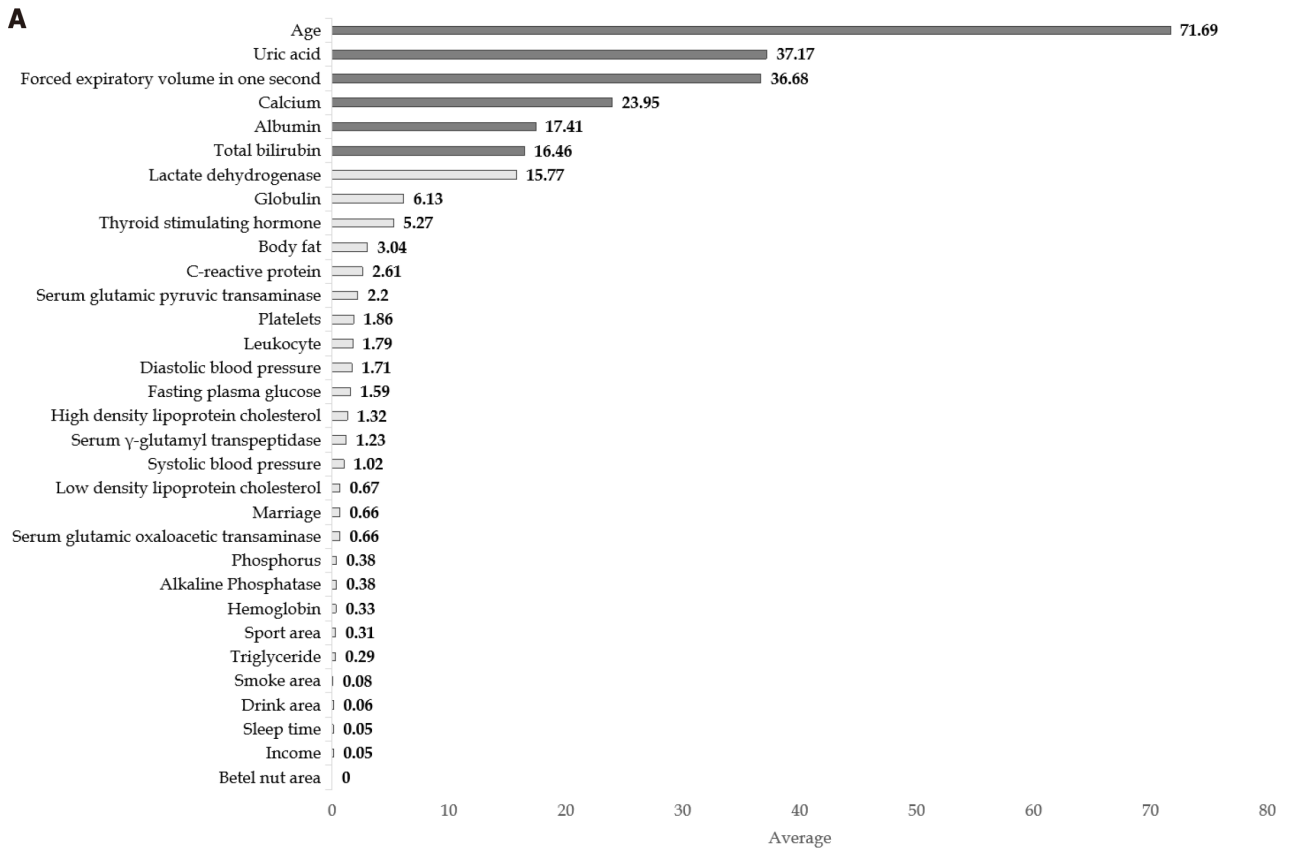
SGB: Stochastic gradient boosting; XGBoost: eXtreme gradient boosting; NAFLD: Non-alcoholic fatty liver.

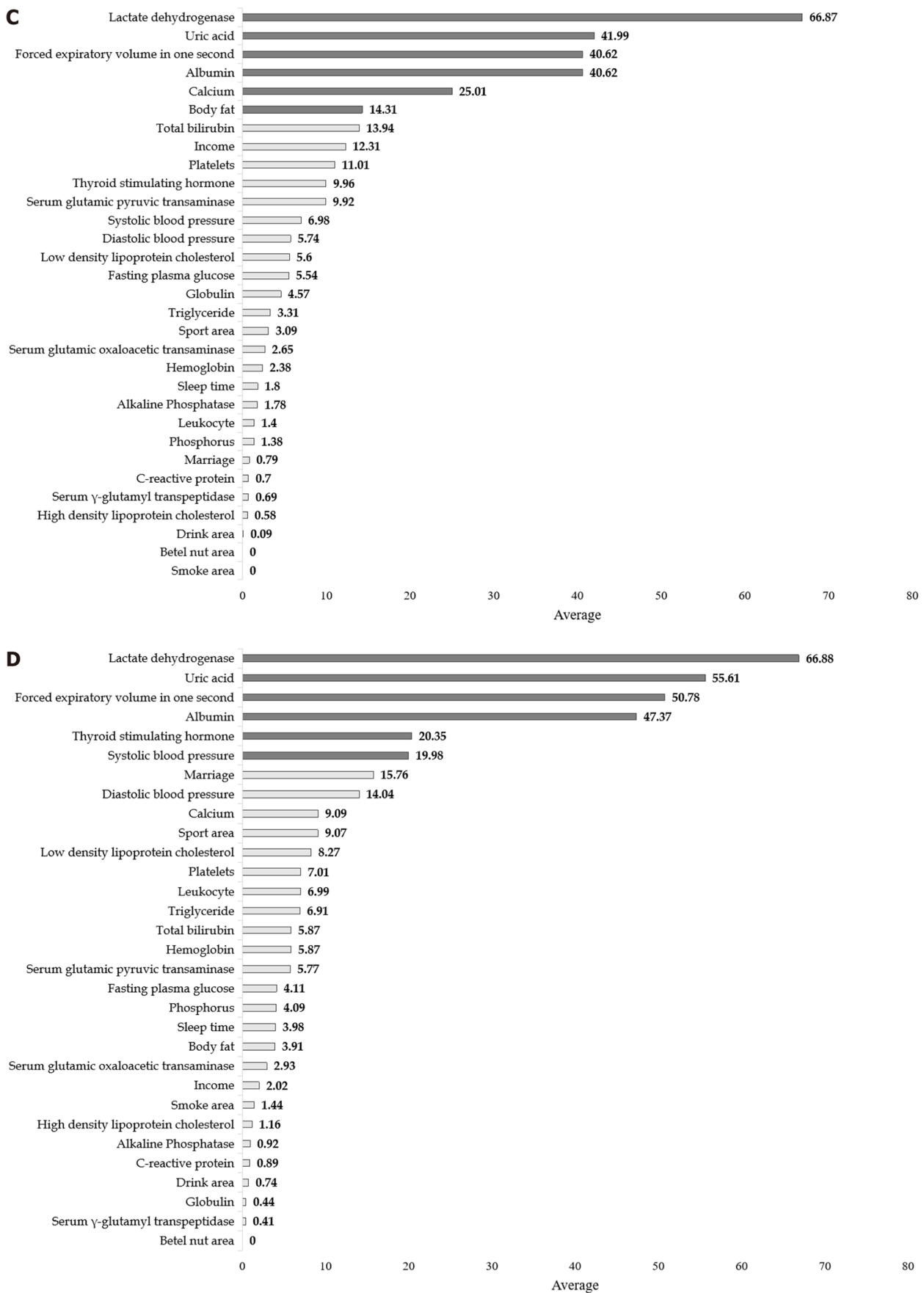
and a similar finding was made for Japanese[37]. Again, our results are consistent with these findings.

In our study, we classified our participants into groups of individuals with and without NAFLD. For both groups, Model 2 showed the same order of the four most important impact factors, diverging thereafter, with TSH and SBP the 5th and 6th ranked factors for NAFLD- as opposed to Ca and BF for NAFLD+. Obesity is one of the most important contributors to NAFLD[38], and thus it is reasonable that body fat levels are a key factor in determining such differences.

Thyroid function influences renal function either systemically, hemodynamically, or directly[39]. Impaired renal function could be noted in both hyper- and hypothyroidism. SBP is negatively related to eGFR[40]. Thus, for the NAFLD- group, our findings are consistent with previous results.

In the NAFLD+ group, both TSH and SBP became less important than Ca and BF. Subjects with lower plasma Ca levels have been found to be more susceptible to renal failure[41]. BF has been proposed as a more precise predictor than BMI for predicting cardiovascular diseases[42]. Increased BF has been shown to be significantly related to inflammation and





**Figure 4** All the orders of the ranks of percentage importance were the same as the results in the tables. A: The average of the risk factors derived from three different machine learning methods (Model 1, NAFLD+ with age); B: The ranks of the risk factors derived from three different machine learning methods (Model 1, NAFLD- with age); C: The ranks of the risk factors derived from three different machine learning methods (Model 2, NAFLD+ without age); D: The ranks of the risk factors derived from three different machine learning methods (Model 2, NAFLD- without age). NAFLD: Non-alcoholic fatty liver.

deterioration of renal function[43]. The increased importance of both Ca and BF indicate that fat levels in the human body impact the relative impact of these factors on renal function. In subjects with higher BF levels, Ca and BF become more important and should be considered differently in terms of their relative influence on renal function.

The present study is subject to certain limitations. First, its cross-sectional nature restricts understanding of causality. A longitudinal design study would provide greater insight into causes and consequences. Second, study participants were restricted to women due to their distinct pathophysiology characteristics. Future work will replicate the study with a male cohort. Finally, the present study only considers ethnic Chinese participants, thus caution should be taken in extrapolating the findings to other ethnic groups.

## CONCLUSION

In conclusion, eGFR levels were lower in the NAFLD+ group compared to the NAFLD- group. Age was found to be the most important single factor affecting eGFR. After removing age, LDH, UA, FEV1 and Alb were four leading impact factors in both groups, followed by TSH and SBP for the NAFLD- group, and Ca and BF for the NAFLD+ group.

## FOOTNOTES

**Author contributions:** Chen IC participated in study design and oversight; Chou LJ participated in study design and data collection; Huang SC participated in data collection and data analysis; Lee SS and Chu TW both participated in data analysis and drafted the manuscript. All authors read and approved the final manuscript.

**Supported by** the Kaohsiung Armed Forces General Hospital.

**Institutional review board statement:** Ethics approval and participant consent: The research plan was reviewed and approved by the Institutional Review Board of Kaohsiung Armed Forces General Hospital on July 1, 2023 prior to the start of the study.

**Informed consent statement:** All the authors consent to the publication.

**Conflict-of-interest statement:** All authors have no conflicts of interest.

**Data sharing statement:** Availability of data and materials: Data available on request due to privacy/ethical restrictions.

**Open-Access:** This article is an 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: <https://creativecommons.org/licenses/by-nc/4.0/>

**Country/Territory of origin:** Taiwan

**ORCID number:** Ta-Wei Chu 0000-0002-7629-7854; Shang-Sen Lee 0000-0003-4103-113X.

**S-Editor:** Liu JH

**L-Editor:** A

**P-Editor:** Zheng XM

## REFERENCES

- 1 **Ansari S**, Haboubi H, Haboubi N. Adult obesity complications: challenges and clinical impact. *Ther Adv Endocrinol Metab* 2020; **11**: 2042018820934955 [PMID: 32612803 DOI: 10.1177/2042018820934955]
- 2 **Liu Q**, Bengmark S, Qu S. The role of hepatic fat accumulation in pathogenesis of non-alcoholic fatty liver disease (NAFLD). *Lipids Health Dis* 2010; **9**: 42 [PMID: 20426802 DOI: 10.1186/1476-511X-9-42]
- 3 **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]
- 4 **Loomba R**, Friedman SL, Shulman GI. Mechanisms and disease consequences of nonalcoholic fatty liver disease. *Cell* 2021; **184**: 2537-2564 [PMID: 33989548 DOI: 10.1016/j.cell.2021.04.015]
- 5 **Targher G**, Chonchol MB, Byrne CD. CKD and nonalcoholic fatty liver disease. *Am J Kidney Dis* 2014; **64**: 638-652 [PMID: 25085644 DOI: 10.1053/j.ajkd.2014.05.019]
- 6 **Kiapidou S**, Liava C, Kalogirou M, Akriviadis E, Sinakos E. Chronic kidney disease in patients with non-alcoholic fatty liver disease: What the Hepatologist should know? *Ann Hepatol* 2020; **19**: 134-144 [PMID: 31606352 DOI: 10.1016/j.aohep.2019.07.013]
- 7 **Yasui K**, Sumida Y, Mori Y, Mitsuyoshi H, Minami M, Itoh Y, Kanemasa K, Matsubara H, Okanou T, Yoshikawa T. Nonalcoholic steatohepatitis and increased risk of chronic kidney disease. *Metabolism* 2011; **60**: 735-739 [PMID: 20817213 DOI: 10.1016/j.metabol.2010.07.022]
- 8 **Chalasan N**, Younossi Z, Lavine JE, Diehl AM, Brunt EM, Cusi K, Charlton M, Sanyal AJ; American Gastroenterological Association;

- American Association for the Study of Liver Diseases; American College of Gastroenterology. The diagnosis and management of non-alcoholic fatty liver disease: practice guideline by the American Gastroenterological Association, American Association for the Study of Liver Diseases, and American College of Gastroenterology. *Gastroenterology* 2012; **142**: 1592-1609 [PMID: 22656328 DOI: 10.1053/j.gastro.2012.04.001]
- 9 **Cao Y**, Du Y, Jia W, Ding J, Yuan J, Zhang H, Zhang X, Tao K, Yang Z. Identification of biomarkers for the diagnosis of chronic kidney disease (CKD) with non-alcoholic fatty liver disease (NAFLD) by bioinformatics analysis and machine learning. *Front Endocrinol (Lausanne)* 2023; **14**: 1125829 [PMID: 36923221 DOI: 10.3389/fendo.2023.1125829]
- 10 **Yuk H**, Gim J, Min JK, Yun J, Heo TY. Artificial Intelligence-based Prediction of Diabetes and Prediabetes Using Health Checkup Data in Korea. *Appl Artif Intell* 2022; **36** [DOI: 10.1080/08839514.2022.2145644]
- 11 **Wu X**, Tsai SP, Tsao CK, Chiu ML, Tsai MK, Lu PJ, Lee JH, Chen CH, Wen C, Chang SS, Hsu CY, Wen CP. Cohort Profile: The Taiwan MJ Cohort: half a million Chinese with repeated health surveillance data. *Int J Epidemiol* 2017; **46**: 1744-1744g [PMID: 28204597 DOI: 10.1093/ije/dyw282]
- 12 **Foundation MHR**. The introduction of MJ health database. MJ Health Research Foundation Technical Report, MJHRF-TR-01. 2016
- 13 **Wu CZ**, Huang LY, Chen FY, Kuo CH, Yeih DF. Using Machine Learning to Predict Abnormal Carotid Intima-Media Thickness in Type 2 Diabetes. *Diagnostics (Basel)* 2023; **13** [PMID: 37296685 DOI: 10.3390/diagnostics13111834]
- 14 **Friedman JH**. Greedy function approximation: A gradient boosting machine. *Ann Stat* 2001; **29**: 1189-1232 [DOI: 10.1214/aos/1013203451]
- 15 **Chen T**, Guestrin C. XGBoost: A Scalable Tree Boosting System. Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining; San Francisco, California, USA: Association for Computing Machinery; 2016; 785-794 [DOI: 10.1145/2939672.2939785]
- 16 **Torlay L**, Perrone-Bertolotti M, Thomas E, Baci M. Machine learning-XGBoost analysis of language networks to classify patients with epilepsy. *Brain Inform* 2017; **4**: 159-169 [PMID: 28434153 DOI: 10.1007/s40708-017-0065-7]
- 17 **Tay JK**, Narasimhan B, Hastie T. Elastic Net Regularization Paths for All Generalized Linear Models. *J Stat Softw* 2023; **106** [PMID: 37138589 DOI: 10.18637/jss.v106.i01]
- 18 **Weinstein JR**, Anderson S. The aging kidney: physiological changes. *Adv Chronic Kidney Dis* 2010; **17**: 302-307 [PMID: 20610357 DOI: 10.1053/j.ackd.2010.05.002]
- 19 **Certo M**, Tsai CH, Pucino V, Ho PC, Mauro C. Lactate modulation of immune responses in inflammatory versus tumour microenvironments. *Nat Rev Immunol* 2021; **21**: 151-161 [PMID: 32839570 DOI: 10.1038/s41577-020-0406-2]
- 20 **Rabinowitz JD**, Enerbäck S. Lactate: the ugly duckling of energy metabolism. *Nat Metab* 2020; **2**: 566-571 [PMID: 32694798 DOI: 10.1038/s42255-020-0243-4]
- 21 **Nielsen VK**, Kemp E, Laursen T. Lactic dehydrogenase in kidney tissue and renal disease. Adaptive change of the synthesis in acute failure. *Acta Med Scand* 1968; **184**: 109-119 [PMID: 5703958 DOI: 10.1111/j.0954-6820.1968.tb02430.x]
- 22 **Delacour JL**, Floriot C, Wagschal G, Daoudal P, Chambers R, Xuan PB. Non-cardiac pulmonary edema following intravenous contrast injection. *Intensive Care Med* 1988; **15**: 49-50 [PMID: 3230202 DOI: 10.1007/BF00255637]
- 23 **Cirillo P**, Sato W, Reungjui S, Heinig M, Gersch M, Sautin Y, Nakagawa T, Johnson RJ. Uric acid, the metabolic syndrome, and renal disease. *J Am Soc Nephrol* 2006; **17**: S165-S168 [PMID: 17130256 DOI: 10.1681/ASN.2006080909]
- 24 **Nakanishi N**, Okamoto M, Yoshida H, Matsuo Y, Suzuki K, Tatara K. Serum uric acid and risk for development of hypertension and impaired fasting glucose or Type II diabetes in Japanese male office workers. *Eur J Epidemiol* 2003; **18**: 523-530 [PMID: 12908717 DOI: 10.1023/A:1024600905574]
- 25 **Mazzali M**, Hughes J, Kim YG, Jefferson JA, Kang DH, Gordon KL, Lan HY, Kivlighn S, Johnson RJ. Elevated uric acid increases blood pressure in the rat by a novel crystal-independent mechanism. *Hypertension* 2001; **38**: 1101-1106 [PMID: 11711505 DOI: 10.1161/hy1101.092839]
- 26 **Sánchez-Lozada LG**, Lanaspá MA, Cristóbal-García M, García-Arroyo F, Soto V, Cruz-Robles D, Nakagawa T, Yu MA, Kang DH, Johnson RJ. Uric acid-induced endothelial dysfunction is associated with mitochondrial alterations and decreased intracellular ATP concentrations. *Nephron Exp Nephrol* 2012; **121**: e71-e78 [PMID: 23235493 DOI: 10.1159/000345509]
- 27 **Rabadi MM**, Kuo MC, Ghaly T, Rabadi SM, Weber M, Goligorsky MS, Ratliff BB. Interaction between uric acid and HMGB1 translocation and release from endothelial cells. *Am J Physiol Renal Physiol* 2012; **302**: F730-F741 [PMID: 22189943 DOI: 10.1152/ajprenal.00520.2011]
- 28 **Kurts C**. A crystal-clear mechanism of chronic kidney disease. *Kidney Int* 2013; **84**: 859-861 [PMID: 24172728 DOI: 10.1038/ki.2013.251]
- 29 **Corry DB**, Eslami P, Yamamoto K, Nyby MD, Makino H, Tuck ML. Uric acid stimulates vascular smooth muscle cell proliferation and oxidative stress via the vascular renin-angiotensin system. *J Hypertens* 2008; **26**: 269-275 [PMID: 18192841 DOI: 10.1097/HJH.0b013e3282f240bf]
- 30 **David S**, Edwards CW. Forced Expiratory Volume. [Updated 2022 Aug 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023 Jan-. Available from: <https://www.ncbi.nlm.nih.gov/books/NBK540970/>
- 31 **Navaneethan SD**, Mandayam S, Arrigain S, Rahman M, Winkelmayr WC, Schold JD. Obstructive and Restrictive Lung Function Measures and CKD: National Health and Nutrition Examination Survey (NHANES) 2007-2012. *Am J Kidney Dis* 2016; **68**: 414-421 [PMID: 27130720 DOI: 10.1053/j.ajkd.2016.03.415]
- 32 **Yu D**, Chen T, Cai Y, Zhao Z, Simmons D. Association between pulmonary function and renal function: findings from China and Australia. *BMC Nephrol* 2017; **18**: 143 [PMID: 28460631 DOI: 10.1186/s12882-017-0565-y]
- 33 **Lang J**, Scherzer R, Weekley CC, Tien PC, Grunfeld C, Shlipak MG. Serum albumin and short-term risk for mortality and cardiovascular disease among HIV-infected veterans. *AIDS* 2013; **27**: 1339-1343 [PMID: 23343914 DOI: 10.1097/QAD.0b013e32835f1dd6]
- 34 **Goldwasser P**, Feldman J. Association of serum albumin and mortality risk. *J Clin Epidemiol* 1997; **50**: 693-703 [PMID: 9250267 DOI: 10.1016/S0895-4356(97)00015-2]
- 35 **Feldman JG**, Gange SJ, Bacchetti P, Cohen M, Young M, Squires KE, Williams C, Goldwasser P, Anastos K. Serum albumin is a powerful predictor of survival among HIV-1-infected women. *J Acquir Immune Defic Syndr* 2003; **33**: 66-73 [PMID: 12792357 DOI: 10.1097/00126334-200305010-00010]
- 36 **Lang J**, Katz R, Ix JH, Gutierrez OM, Peralta CA, Parikh CR, Satterfield S, Petrovic S, Devarajan P, Bennett M, Fried LF, Cummings SR, Sarnak MJ, Shlipak MG. Association of serum albumin levels with kidney function decline and incident chronic kidney disease in elders. *Nephrol Dial Transplant* 2018; **33**: 986-992 [PMID: 28992097 DOI: 10.1093/ndt/gfx229]
- 37 **Cheng T**, Wang X, Han Y, Hao J, Hu H. The level of serum albumin is associated with renal prognosis and renal function decline in patients with chronic kidney disease. *BMC Nephrol* 2023; **24**: 57 [PMID: 36922779 DOI: 10.1186/s12882-023-03110-8]

- 38 National Guideline C. National Institute for Health and Care Excellence: Guidelines. Non-Alcoholic Fatty Liver Disease: Assessment and Management. London: National Institute for Health and Care Excellence (NICE) Copyright © National Institute for Health and Care Excellence, 2016
- 39 **Iglesias P**, Bajo MA, Selgas R, Díez JJ. Thyroid dysfunction and kidney disease: An update. *Rev Endocr Metab Disord* 2017; **18**: 131-144 [PMID: 27864708 DOI: 10.1007/s11154-016-9395-7]
- 40 **Peralta CA**, Whooley MA, Ix JH, Shlipak MG. Kidney function and systolic blood pressure new insights from cystatin C: data from the Heart and Soul Study. *Am J Hypertens* 2006; **19**: 939-946 [PMID: 16942937 DOI: 10.1016/j.amjhyper.2006.02.007]
- 41 Kidney Failure Risk Factor: Serum Calcium. Available from: <https://www.kidney.org/content/kidney-failure-risk-factor-serum-calcium>.
- 42 **Zeng Q**, Dong SY, Sun XN, Xie J, Cui Y. Percent body fat is a better predictor of cardiovascular risk factors than body mass index. *Braz J Med Biol Res* 2012; **45**: 591-600 [PMID: 22510779 DOI: 10.1590/S0100-879X2012007500059]
- 43 **Chen YY**, Fang WH, Wang CC, Kao TW, Chang YW, Yang HF, Wu CJ, Sun YS, Chen WL. Changes of Percent Body Fat as a Useful Surrogate for Risk of Declined Renal Function. *Sci Rep* 2018; **8**: 17289 [PMID: 30470784 DOI: 10.1038/s41598-018-35601-2]





Published by **Baishideng Publishing Group Inc**  
7041 Koll Center Parkway, Suite 160, Pleasanton, CA 94566, USA  
**Telephone:** +1-925-3991568  
**E-mail:** [office@baishideng.com](mailto:office@baishideng.com)  
**Help Desk:** <https://www.f6publishing.com/helpdesk>  
<https://www.wjgnet.com>

