Computers in Biology and Medicine 135 (2021) 104622

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Application of computer tongue image analysis technology in the diagnosis of NAFLD

Tao Jiang [a,1], Xiao-jing Guo [a,1], Li-ping Tu [a], Zhou Lu [a], Ji Cui [a], Xu-xiang Ma [a], Xiao-juan Hu [a], Xing-hua Yao [a], Long-tao Cui [a], Yong-zhi Li [b], Jing-bin Huang [a,*], Jia-tuo Xu [a,**]

[a] Basic Medical College, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Shanghai, 201203, China
[b] China Astronaut Training Center, Beijing, 100084, China

ARTICLE INFO

Keywords:
NAFLD
Tongue image
Diagnostic model
Computer intelligent technology
Deep learning

ABSTRACT

Nonalcoholic fatty liver disease (NAFLD), a leading cause of chronic hepatic disease, can progress to liver fibrosis, cirrhosis, and hepatocellular carcinoma. Therefore, it is extremely important to explore early diagnosis and screening methods. In this study, we developed models based on computer tongue image analysis technology to observe the tongue characteristics of 1778 participants (831 cases of NAFLD and 947 cases of non-NAFLD). Combining quantitative tongue image features, basic information, and serological indexes, including the hepatic steatosis index (HSI) and fatty liver index (FLI), we utilized machine learning methods, including Logistic Regression, Support Vector Machine (SVM), Random Forest (RF), Gradient Boosting Decision Tree (GBDT), Adaptive Boosting Algorithm (AdaBoost), Naïve Bayes, and Neural Network for NAFLD diagnosis. The best fusion model for diagnosing NAFLD by Logistic Regression, which contained the tongue image parameters, waist circumference, BMI, GGT, TG, and ALT/AST, achieved an AUC of 0.897 (95% CI, 0.882–0.911), an accuracy of 81.70% with a sensitivity of 77.62% and a specificity of 85.22%; in addition, the positive likelihood ratio and negative likelihood ratio were 5.25 and 0.26, respectively. The application of computer intelligent tongue diagnosis technology can improve the accuracy of NAFLD diagnosis and may provide a convenient technical reference for the establishment of early screening methods for NAFLD, which is worth further research and verification.

1. Introduction

Nonalcoholic fatty liver disease (NAFLD) has a global prevalence of approximately 25%, showing a younger trend that is continuing to rise. NAFLD is the leading cause of chronic hepatic disease and poses a huge clinical and economic burden to all societies [1–3]. NAFLD is a chronic progressive disease and includes non-alcoholic fatty liver (NAFL) and non-alcoholic steatohepatitis (NASH) [1,4]. The latter is closely related to liver fibrosis, cirrhosis, and hepatocellular carcinoma [5], and 10%–29% of NASH patients develop cirrhosis within 10 years [6].

NAFLD has also been found to be a risk factor for extrahepatic diseases such as subclinical atherosclerosis, decreased bone mineral density, renal insufficiency, obstructive sleep apnea, and polycystic ovary syndrome [7]. Studies on the mortality rates of NAFLD have demonstrated that even in patients with simple steatosis, the mortality risk increases by 71%, whereas with an aggravation of the fatty liver, the mortality risk of NASH, liver fibrosis, and cirrhosis increase by 114%, 144%, and 279%, respectively [8]. Early lifestyle interventions, such as weight loss, play an important role in reversing and slowing down the progression of nonalcoholic steatohepatitis [9,10]. Therefore, it is worth developing methods for the early diagnosis of NAFLD, which may assist in providing an ideal therapeutic option for early life intervention, thus slowing down and reversing the transformation of NAFL into NASH, reducing NAFLD-related mortality, and reducing the economic burden of global health.

NAFLD is a pathological condition characterized by excessive lipid accumulation in the liver and is defined by the presence of steatosis in >5% of hepatocytes according to a histological analysis. As a first-line diagnostic technique in patients with NAFLD, ultrasonography can be used to robustly diagnose moderate or severe steatosis but is operator-

* Corresponding author.
** Corresponding author.
E-mail addresses: jiangtao@shutcm.edu.cn (T. Jiang), sunflowergxj@126.com (X.-j. Guo), jmy61@163.com (J.-b. Huang), xjt@fudan.edu.cn (J.-t. Xu).
[1] Tao Jiang and Xiao-jing Guo are co-first authors.

https://doi.org/10.1016/j.compbiomed.2021.104622
Received 22 March 2021; Received in revised form 28 June 2021; Accepted 28 June 2021
Available online 1 July 2021

dependent and has limited sensitivity. When steatosis is less than 20% or when the body mass index (BMI) is $> 40\ kg/m^2$, ultrasonography cannot reliably detect hepatic steatosis. In addition, during large-scale screening, the implementation capability and cost of ultrasonography significantly affect its feasibility [5]. It is therefore necessary to develop a more convenient and accurate screening tool for NAFLD. The fatty liver index (FLI) and hepatic steatosis index (HSI), based on biochemical and anthropometric parameters, have been reported to be used in the diagnosis and prediction of NAFLD [11–14]. However, optimization of the diagnostic accuracy and the specificity of NAFLD remain to be further studied.

A tongue inspection used in traditional Chinese medicine (TCM), a simple and intuitive method for obtaining useful clinical information, is regarded as a non-invasive technique for the identification of diseases and their symptoms, determination of disease severity, and an evaluation of the efficacy of different treatment methods [15–19]. However, a TCM tongue diagnosis relies on the visual observation of Chinese physicians, including the color, texture, and shape of the tongue, and lacks objective quantitative evidence, which limits the clinical application of tongue diagnosis technology to a certain extent [20,21].

With the development of computer imaging technology, digital image processing technology has been widely used in intelligent tongue diagnosis [21,22]. Image segmentation [23,24], image correction [25], and image analysis technologies applied to the tongue image analysis [26,27] have provided convenient conditions for the application of an objective tongue diagnosis. Studies have shown that a tongue diagnosis serves as a screening tool for T2DM [28–30] and early breast cancer [31], and plays a predictive role in metabolic syndrome [32]. However, TCM tongue diagnosis technology applied to the diagnosis and prediction of NAFLD has yet to be reported. Therefore, we assume that, based on the current NAFLD diagnosis model, the addition of a computer intelligent tongue image diagnosis technology can obtain a higher accuracy and predictive value, and can become a convenient and economical method for the early screening of NAFLD.

2. Material and methods

2.1. Participants

We recruited volunteers from the Physical Examination Center of Shuguang Hospital, which is affiliated with Shanghai University of TCM, from July 2018 to December 2018, and a total of 2115 subjects participated in the study. The diagnostic criteria for NAFLD refer to the "Guidelines of Prevention and Treatment for Nonalcoholic Fatty Liver Disease: 2018 Update" [33].

The inclusion criteria of this study were as follows: between 25 and 80 years in age, fatty liver diagnosis through ultrasound, and no treatment received. We excluded patients with malignant tumors, metabolic liver disease, or autoimmune liver disease; weekly alcohol consumption of more than 140 g for women and 210 g for men during the past 12 months; those diagnosed with NAFLD caused by a known extrahepatic disease (i.e., hypothyroidism, Cushing's syndrome, or polycystic ovary syndrome) and/or drug-related factors (i.e., long-term use of glucocorticoids and tamoxifen [34]); women during pregnancy or breastfeeding; and those who failed to cooperate with the researchers. At the same time, those who had no NAFLD diagnosis, no history of liver disease, and normal liver function were included as the control group for analysis.

2.2. Tongue image and clinical data collection

All participants were required to avoid greasy food and fast after 10 p.m. the day before the observation. Routine biological blood samples were tested by the Department of Laboratory Medicine, Shuguang Hospital, using the standard protocols measured with an automatic Beckman Coulter AU5800 biochemical analyzer (Beckman Coulter Inc. Brea, CA, USA), including liver and renal function indexes, blood lipids, and blood glucose indexes. An ultrasound examination was conducted by a sonographer with more than 10 years of work experience using a FibroScan-520 (Echosens, Paris, France). The researchers collected basic information (Table 1) and collected tongue images under fasting conditions. The collection time was 8:00–10:00 a.m., and the research process and collected clinical information are as shown in Fig. 1 and Table 1.

Tongue images were collected using a TFDA-1 tongue diagnosis instrument, which was developed by the Chinese Medical Diagnostics Laboratory of Shanghai University of TCM (Fig. 2A and B). The instrument was composed of CCD equipment and a standard D50 light source (main parameters = color rendering index of 97 and color temperature of 5003 K), providing stable and standard acquisition conditions [35], which had also been used for diabetes and other diseases [36]. The tongue image acquisition process is as follows: ① Instruct the subjects to stay calm, avoid mood swings or strenuous exercise, and clean their mouth and tongue of any food residue, foreign objects, or stains. ② Sterilize the contact part between the outer ring of the "TFDA-1" desktop tongue examination instrument and the face, and check and set the shooting parameters of the tongue examination instrument. ③ Instruct the subjects to sit upright, look up at the lens of the tongue examination instrument, and close their eyes tightly. After the light source of the instrument was sufficiently stable, the subjects were instructed to stretch out their tongue naturally and keep it relaxed during the TFDA-1 tongue examination. After observing the tongue image in the preview window of the instrument, the shutter was pressed to complete the image capture. If the acquisition needed to be repeated, the subjects were allowed to rest for 3–5 min (Fig. 2C and D).

2.3. Tongue image feature extraction

The tongue images include color, shape, and texture features. We used the TCM Tongue Image Diagnostic Analysis System (TDAS) V1.0 (software copyright registration No. 2018SR033451) to automatically extract the tongue image features [20].

First, to control the noise of the face and background areas around the tongue body region, all available raw tongue images were isolated and cropped manually to select the region of interest (ROI) of the tongue image, and then normalized to the same size (400 pixels × 400 pixels) as the tongue region.

Second, in this research, the tongue body region was segmented using the Mask R–CNN framework, and the Faster R–CNN framework [37–39] was used for the tongue shape and texture recognition. Mask R–CNN and Faster R–CNN are current mainstream object detection

Table 1
Basic information and clinical content collected from the subjects.

Basic information	Name, sex, and age
Physical examinations	Height, weight, waist circumference (WC), hip circumference, systolic blood pressure (SBP), diastolic blood pressure (DBP), body mass index (BMI), ratio of waist circumference to hip circumference (WHR), ratio of waist circumference to height (WHtR)
Serological indexes	Related indexes of liver function: alanine aminotransferase (ALT), aspartate aminotransferase (AST), gamma glutamyl aminotransferase (GGT), alkaline phosphatase (ALP) Blood lipid related indexes: total cholesterol (TC), triglyceride (TG), low-density lipoprotein cholesterol (LDL-C), and high-density lipoprotein cholesterol (HDL-C) Blood glucose related indexes: fasting blood glucose (FPG) and glycosylated hemoglobin (HbA_{1C}) Related indexes of renal function: serum uric acid (UA), urea nitrogen (BUN), and creatinine (SCr)
Diseases and drugs	Acute or chronic diseases other than NAFLD and the drugs used.
Drinking	Daily alcohol consumption, drinking duration, and calculation of daily alcohol intake.

Note: BMI is the weight divided by the square of the height.

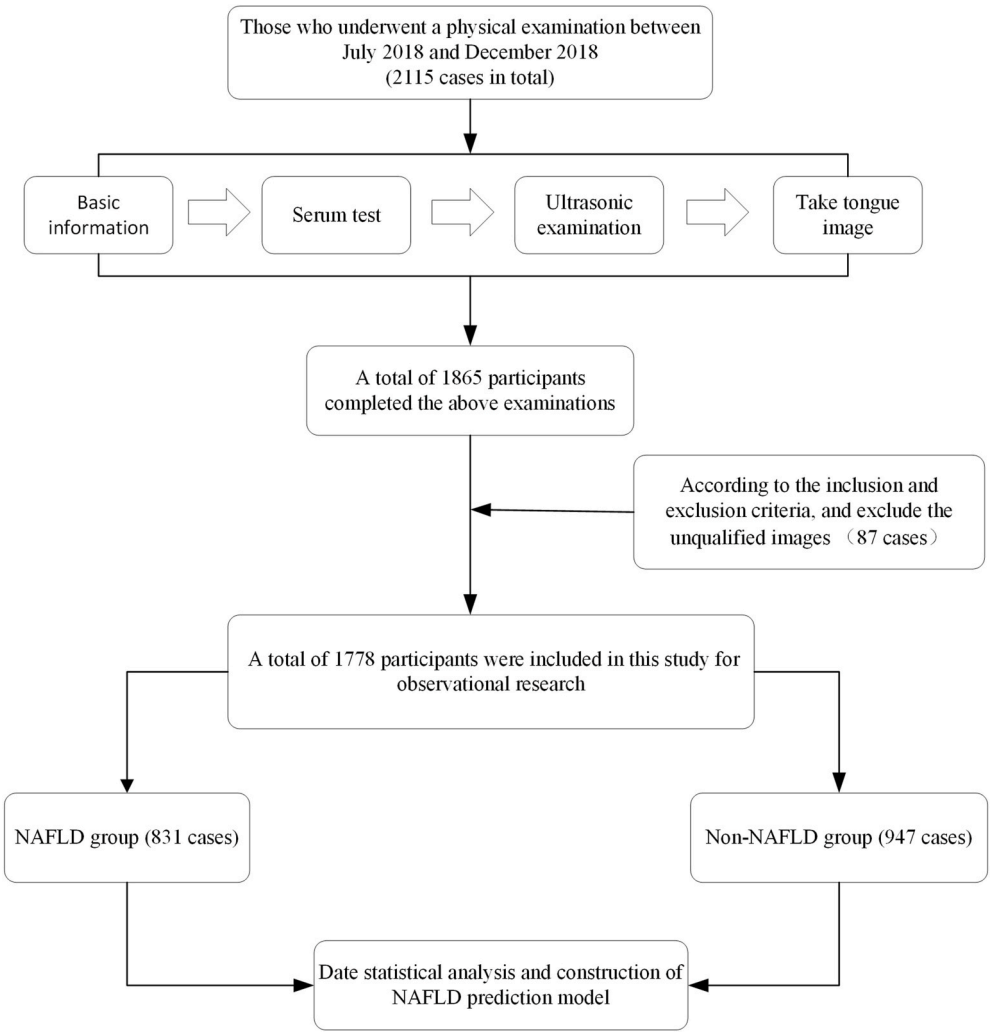

Fig. 1. Entire research processes.

neural networks. Based on the Faster-RCNN framework, the Mask-RCNN framework replaces ROI pooling with ROI align and realizes the precise location of the bounding box through a bilinear interpolation method with a mask predictor based on a box regression. In this study, CNN models were trained using the PyTorch (version 1.3.1) and Python (version 3.6) frameworks on the Ubuntu system (version 16.04). Models were developed on DEVTOP AIX4750 hardware produced by OMNISKY with an NVIDIA GTX 1080Ti GPU, i7-6850 K CPU, 64 GB of DDR4 RAM, and 512 GB SDD [40].

For automatic tongue body segmentation using Mask R–CNN, we annotated 1100 raw tongue images, among which 1000 tongue images were used to train the Mask R–CNN model, and 100 tongue images were used to test the model. The mean of the average precision (mAP) and mean of intersection over union (mIoU) across the tongue images in the test set were used to evaluate the mask R–CNN model. Our model achieved an mAP of 99.85% and an mIoU of 95.04% in the test set. The accuracy of the Mask R–CNN model training was close to 100%. The specific segmentation effect on the test set of 100 tongue images is shown in Fig. 3. The inference time of the training model for tongue image segmentation in this study was 0.443–0.599 s, the mIoU was 95.06%, and mPA was 99.85%. This shows that the automatic recognition and segmentation model of the tongue image region established by Mask R–CNN achieved a pixel-level accuracy.

The texture and shape features of the tongue images used in this study included a fissured tongue, tooth marked tongue, stasis tongue, spotted tongue, greasy coating, peeled coating, and rotten coating. The faster R–CNN model was used to identify the shape and texture of the tongue. The tongue image dataset constructed by experts consisted of 1962 fissured tongue images, 1732 tooth marked tongue images, 1383 stasis tongue images, 932 spotted tongue images, 1949 greasy coating images, 598 peeled coating images, and 120 rotten coating images. In each tongue image, regions of interest were marked manually using LabelImg software according to the annotations of the experts. The tongue images in the dataset were partitioned into three sets, including the training, validation, and testing sets, i.e., 80% for training, 10% for validation, and 10% for testing. The overall accuracy of the model was over 90%, and the faster R–CNN model was able to complete the target detection of multi-labeled tongue shapes and texture features (Fig. 4), demonstrating their specific parts with an excellent visualization effect.

The transfer learning strategy of deep learning was used to reload the pre-trained parameters obtained by training the model on a large dataset into a network structure, and thus we can train the R–CNN model more effectively and quickly and avoid an overfitting. The parameters were initialized using weights pretrained on the ImageNet dataset. The initial parameter settings were 40,000 iterations, a learning rate of 0.001, a stochastic gradient descent optimization, a momentum of 0.9, a weight decay of 0.0001, and a batch size of 128. The multi-task loss function of Mask R–CNN combines the losses of classification, localization, and a segmentation mask, where the losses of the classification and the bounding box regression are the same as those in Faster R–CNN. We then fine-tuned the parameters based on our tongue dataset and finally realized a qualitative analysis of the tongue shape and texture.

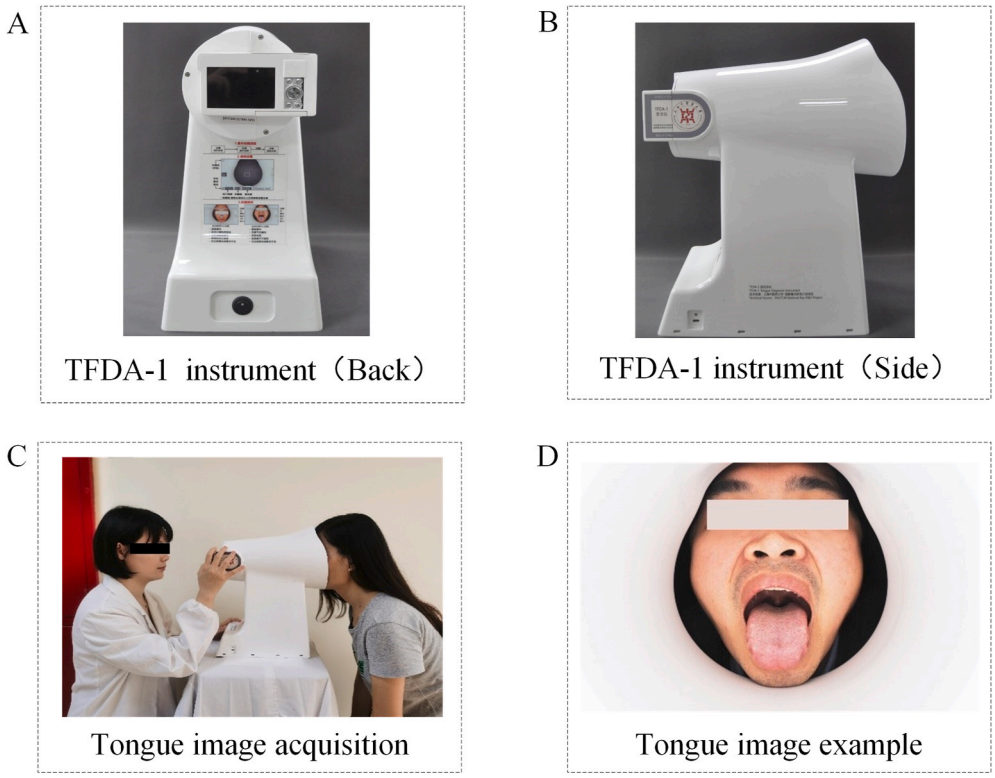

Fig. 2. Tongue image acquisition process. A and B show the appearance of TFDA-1. C is a schematic diagram of the TFDA-1 operation. D is an image of the subject's tongue extension.

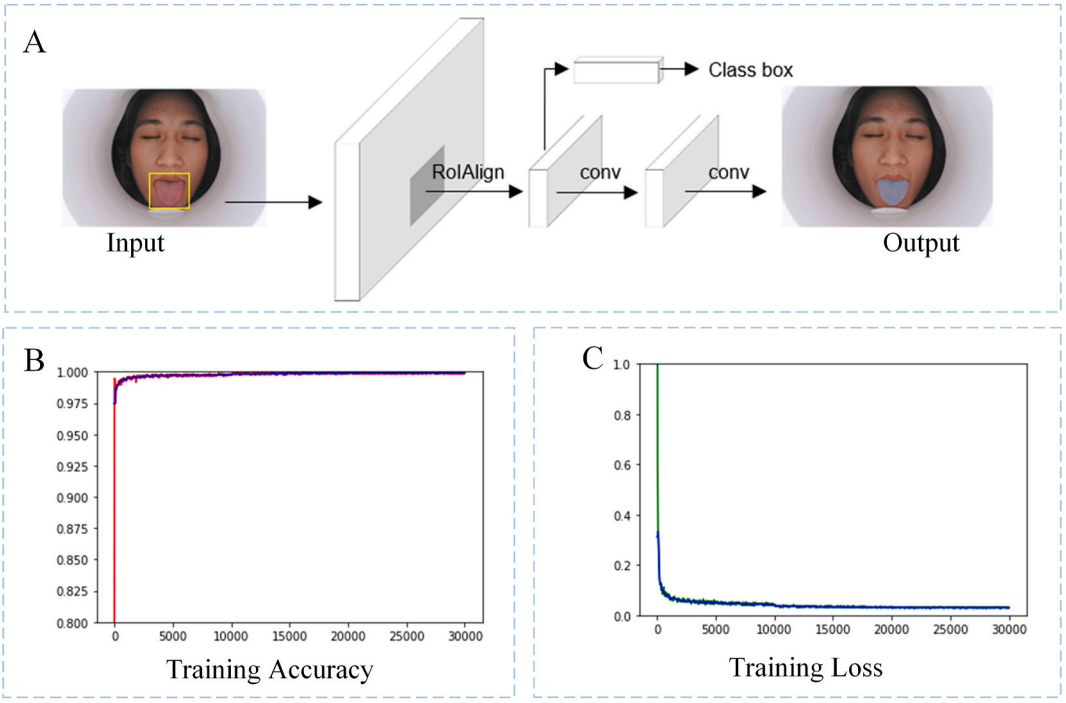

Fig. 3. Framework of Mask R–CNN. A is the Mask R–CNN framework, and B and C are the changes in training accuracy and loss of Mask R–CNN, respectively.

Third, the regional segmentation of the tongue body and tongue coating was accomplished through the split and merge algorithm and color threshold method [29,30,35,36,41]. The tongue image color feature index based on the $L^*a^*b^*$ color space is the conversion result of the RGB color space value [42]. The tongue coating index perAll was calculated as a ratio of the tongue coating pixels to the total tongue body pixels. Finally, the color parameters, texture indicators, and tongue coating index were successively extracted [41]. The main indicators are the tongue body (TB) color feature indicators (including TB-L, TB-a, and TB-b) and tongue coating (TC) color feature indicators (including TC-L,

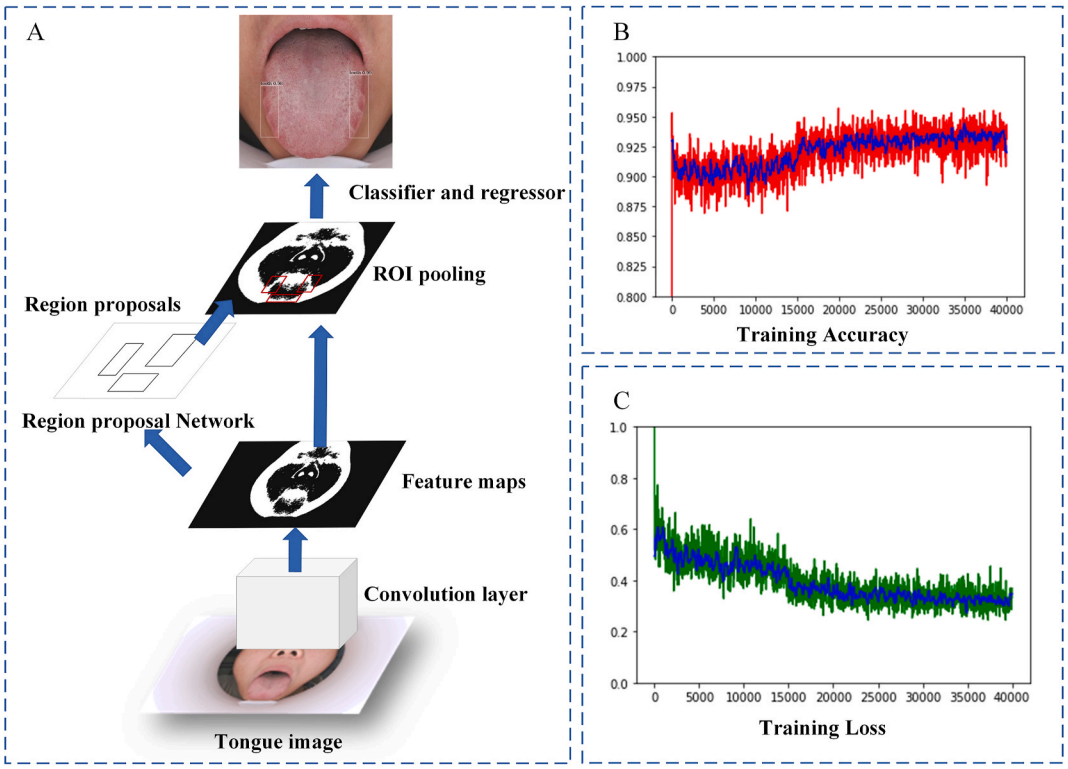

Fig. 4. Faster R–CNN training optimization process.

TC-a, and TC-b). Fig. 5 shows the process of extracting the color, shape, and texture features of the tongue using TDAS.

2.4. Data analysis

Statistical analyses were conducted using SPSS Statistics (version 25.0; IBM Co., Armonk, NY, USA). Continuous variables are presented as the median (interquartile range) and categorical variables as numbers (percentages). The Kolmogorov–Smirnov test was used to determine the data normality. Differences between the two groups were evaluated through a Mann–Whitney *U* test in the case of a skewed distribution, and a Kruskal–Wallis test was used for continuous variables with multiple groups of skewed distributions. A Pearson's chi-square test or Fisher's exact test was used for the categorical variables. At the same time, Pearson or Spearman methods were used for the correlation analysis. In addition, we selected C*ohen's d* to evaluate the effect size, and conducted *Cohen's d* calculations on a website (https://www.psychometrica.de/effect_size.html). According to the suggestions for interpreting the magnitude of the effect sizes, when the *Cohen's d* value is less than 0.2, there is no effect; when the *Cohen's d* value is larger or equal to 0.2 and

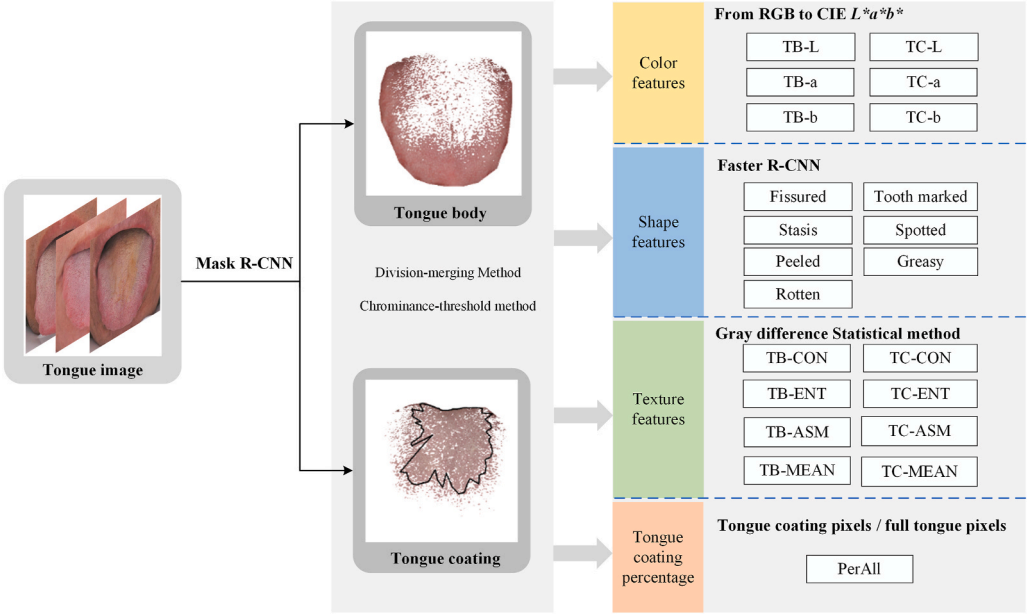

Fig. 5. Automatic tongue image analysis system framework of TDAS.

less than 0.5, the effect is small; when the *Cohen's d* value is larger or equal to 0.5, there is an intermediate effect; and when the *Cohen's d* value is larger or equal to 0.8, the effect is large [43].

Logistic regression, a method used to analyze the relationship between binary or classification results and multiple influencing factors, has been widely applied in multivariate modeling and quantitative studies of TCM syndromes [44,45]. In this study, we used a binary logistic regression to construct the model. The dependent variable was defined as NAFLD. Age and sex were taken as the first level of the dependent variable, and other indicators were taken as the input of the dependent variable as the second level. The forward stepwise (likelihood ratio) method was used, in which the confidence interval for exp (B) was 95%, and the Probability for Stepwise was set as follows: an entry of 0.05, classification cutoff of 0.5, maximum number of iterations of 20, and a constant included in the model.

The receiver operating characteristic (ROC) curves of different models were drawn through a logistic regression, the area under the ROC curve (AUROC) was calculated to evaluate the diagnostic performance of the models, and MedCalc (v19.5.6) was employed to compare the AUROC and confirm the best NAFLD prediction model. The cutoff values of the best NAFLD diagnostic model were calculated according to the maximum principle of the Youden Index [46], and the statistical significance was set at $P < 0.05$.

The evaluation indexes of the model used in this study included the accuracy, sensitivity/recall, specificity, precision, positive likelihood ratio (LR), negative likelihood ratio (NLR), and F1-score, which can be calculated as true positive (TP), true negative (TN), false positive (FP), and false negative (FN), respectively. The specific formula is referred to in the relevant literature [47], (Eqs. (1)– (7)). Meanwhile, according to the NAFLD diagnosis model reported in the literature, the FLI and HSI scores [11,12] were calculated.

Because a logistic regression is a linear regression model, we added a non-linear model method to further analyze the data in this study. We compared with common nonlinear models, including Adaptive Boosting Algorithm (AdaBoost), Gradient Boosting Decision Tree (GBDT), Naïve Bayes, Neural Network, Random Forest, and SVM. We introduced the indicators in the best NAFLD diagnostic model obtained through the logistic regression method in this study into these nonlinear models for calculation. Based on the above evaluation indexes of the model, we compared the results of different modeling methods to determine the best modeling method for the research data.

3. Results

3.1. Differences between NAFLD patients and non-NAFLD participants

A total of 2115 participants participated in this study. Among them, 250 cases were excluded owing to an incomplete collection of clinical data, and 87 cases whose tongue image quality was unqualified (incomplete or missing tongue images) were eliminated. Finally, 1778 cases (831 and 947 cases in the NAFLD and non-NAFLD groups, respectively) were included in this study (Fig. 1). The results showed no statistical difference based on sex or age between the two groups. In the basic physiological indexes, the BMI and WHR of the NAFLD group were higher than those of the control group. Among the blood test indexes, ALT, AST, GGT, SBP, DBP, FPG, TC, TG, and UA in the NAFLD group were higher, whereas high-density lipoprotein was lower than that in the non-NAFLD group ($P < 0.05$). In addition, more participants in the NAFLD group had a metabolic syndrome [33] than those in the non-NAFLD group (37.8% versus 7.9%, $P < 0.05$) (Table 2). Our results show that most of the *Cohen's d* values of the indicators that are statistically different between the NAFLD and control groups are above 0.5, indicating that these differences have an intermediate effect.

Table 2
Differences between non-NAFLD and NAFLD groups.

	Non-NAFLD (n = 947)	NAFLD (n = 831)	*t/Z*	*Cohen's d*	*P*
Age (years)	38 (33–46)	39 (33–47)	−1.282	0.061	0.200
Sex (male/female)	724/223	612/219	1.865	0.065	0.172
Height (cm)	170.39 (8.02)	171.02 (8.41)	−1.634	0.077	0.102
Weight (kg)	68 (61–75)	78 (71–86)	−18.548	0.979	<0.001
BMI (kg/m2)	23.3 (21.5–25.1)	26.5 (24.8–28.4)	−23.344	0.306	<0.001
Waist (cm)	81.93 (8.71)	90.90 (9.01)	−21.316	1.013	<0.001
Hipline (cm)	96 (92–100)	100 (96–105)	−16.325	0.844	<0.001
WHR	0.86 (0.82–0.9)	0.90 (0.87–0.94)	−14.052	0.712	<0.001
WHtR	0.48 (0.45–0.52)	0.53 (0.5–0.56)	−19.639	1.053	<0.001
Total protein (g/L)	75.78 (4.06)	76.42 (3.88)	−3.114	0.161	0.002
Globulin (g/L)	29.84 (3.32)	30.32 (3.5)	−2.720	0.141	0.007
Albumin (g/L)	45.97 (2.63)	46.10 (2.36)	−1.001	4.04	0.317
Prealbumin (g/L)	282.50 (49.61)	298.26 (56.46)	−5.734	0.298	<0.001
SBP (mmHg)	121 (111–133)	129 (119–140)	−10.133	0.495	<0.001
DBP (mmHg)	76 (69–84)	82 (74–89)	−10.436	0.511	<0.001
ALT (U/L)	19 (14–25)	31 (21–47)	−18.795	0.995	<0.001
AST (U/L)	21 (18–24)	24 (20–30)	−11.541	0.568	<0.001
GGT (U/L)	21 (16–28)	32 (23–44)	−17.187	0.892	<0.001
ALP (U/L)	71 (62–85)	77 (66–91)	−5.711	0.5	<0.001
FPG (mmol/L)	4.9 (4.6–5.1)	5.0 (4.8–5.4)	−8.557	0.413	<0.001
TG (mmol/L)	1.13 (0.83–1.49)	1.83 (1.3–2.56)	−18.824	0.998	<0.001
TC (mmol/L)	5.03 (4.48–5.61)	5.31 (4.76–5.97)	−6.977	0.336	<0.001
LDL-C (mmol/L)	2.94 (2.52–3.44)	3.17 (2.7–3.72)	−6.377	0.306	<0.001
HDL-C (mmol/L)	1.33 (1.15–1.56)	1.17 (1.04–1.33)	−12.830	0.639	<0.001
UA (umol/L)	366 (312–417)	405 (347–463)	−10.158	0.5	<0.001
Complicated with metabolic syndrome (%)	7.9%	37.8%	230.986	0.762	<0.001

3.2. Tongue image features of NAFLD patients based on computer imaging technology

Using computer imaging technology, we compared the shape, texture, and color characteristics of the tongue images between the NAFLD and non-NAFLD groups, and found that the proportion of spotted tongue in the NAFLD group was lower than that in the control group ($P < 0.05$), and there were no significant differences in the other tongue features (Table 3). The difference in color characteristic parameters of the tongue images between the two groups was analyzed using a machine learning method. It was found that TB-b and TC-b of the NAFLD group were lower than those of the control group, and the perAll value was higher than that of the control group ($P < 0.05$); the effect size was

Table 3
Differences of tongue shape and texture characteristics between two groups.

	Non-NAFLD (n = 947)	NAFLD (n = 831)	χ^2	*P*
Fissured	439 (46.4%)	390 (46.9%)	0.059	0.809
Tooth Marks	341 (36.0%)	331 (39.8%)	2.752	0.097
Spotted	166 (17.5%)	106 (12.8%)	7.783	0.005
Petechiae	61 (6.4%)	70 (8.4%)	2.548	0.110
Greasy Coating	275 (29.0%)	266 (32.0%)	1.845	0.174
Peeling Coating	40 (4.2%)	34 (4.1%)	0.019	0.889
Rotting Coating	17 (1.8%)	13 (1.6%)	0.142	0.706

small, and there were no significant differences in other indicators (Table 4).

3.3. Correlation analysis of tongue image parameters and physiological indexes

In this study, we analyzed the correlations between serological indexes, basic status of the individual, and tongue image parameters in 831 patients with NAFLD, and calculated the HSI and FLI values of all subjects and conducted a correlation analysis using objective tongue image parameters. The results showed that most of the physiological indexes were related to the b^* value of the tongue coating and tongue body, in which the correlation coefficient between TB-b and sex was the highest, at 0.286 ($P < 0.01$) (Table 5 and Fig. 6).

3.4. NAFLD fusion diagnosis model based on objective tongue image parameters

The main purpose of this study was to explore and construct a convenient and accurate NAFLD diagnosis model based on objective tongue image indicators. To clarify the contributions of the objective tongue image indicators in the diagnosis of NAFLD, we compared the NAFLD diagnostic models with different indicators. Model 1 mainly included basic physiological indices, including sex, age, waist circumference, and BMI. Meanwhile, we added objective tongue image parameters, i.e., tongue texture indexes (perAll and Spotted tongue) and L*a*b values of the tongue body and coating, to model 1 as model 2. Previous studies verified that the HSI and FLI indexes had a high accuracy and specificity in NAFLD screening [14,48]. Therefore, on the basis of model 1, the serological index related to the HSI and FLI indexes, i.e., TG, GGT, and ALT/AST, were included to build model 3, and the same objective tongue image indicators as in model 2 were included in model 3 to build model 4 (Table 6).

The results showed that the objective tongue image parameters played a role in improving the accuracy of different NAFLD diagnostic models (Table 7). Compared with model 1, the inclusion of tongue image parameters TB-b and TC-L in model 2 increased the accuracy of the NAFLD diagnosis from 75.2% to 76.3%, the LR increased from 3.37 to 3.62, the NLR decreased from 0.37 to 0.35, and the classification performance of the two models for NAFLD diagnosis was statistically different ($P < 0.05$). Compared with model 3, the inclusion of tongue image parameters, i.e., TB-L, TB-b, and TC-L, in model 4 increased the accuracy of the NAFLD diagnosis from 80.9% to 81.7%, the LR increased from 4.93 to 5.25, and the NLR was reduced from 0.27 to 0.26, which improved the NAFLD diagnostic model, the difference of which was statistically significant ($P < 0.05$) (Table 8 and Fig. 7).

Model 4, which combined basic indicators, biochemical indicators, and tongue image parameters, obtained the highest LR and the lowest NLR, and was the best diagnostic model for NAFLD in this study. The Youden index of model 4 was 0.6319 (95% CI = 0.5894–0.6595) with a sensitivity of 79.66% and a specificity of 83.53%; that is, when the test result was greater than 0.6319 based on model 4, the tester would be diagnosed with NAFLD.

To further clarify the diagnostic efficacy of the NAFLD diagnostic model fused with the objective tongue image parameters determined in this study, we compared the differences between Model 4 and the diagnostic HSI and FLI models. The accuracies of FLI and HSI in this study were 78.7% and 79.1%, respectively (Tables 9 and 10). The LR and NLR of the HSI diagnostic model were 4.30 and 0.30, respectively, whereas those of the FLI diagnostic model were 4.29 and 0.31, respectively. A further comparative analysis of AUROC found that there were no statistically significant differences in this study (Table 10). However, the accuracy of model 4 was higher than that of the HSI and FLI models, and there were obvious statistical differences, suggesting that the objective tongue image parameters contributed to the diagnosis of NAFLD by a certain degree (Table 11 and Fig. 8).

3.5. Comparison of different modeling methods in classification of NAFLD

The results showed that the Neural Network model achieved the highest accuracy of 80.76%, sensitivity of 76.77%, and specificity of 84.27%, followed by GBDT with an accuracy of 79.47%, random forest with an accuracy of 77.56%, Naïve Bayes with an accuracy of 76.32%, and SVM with an accuracy of 78.07%. The performance of the AdaBoost classification was the worst, with an accuracy of 72.22%. The experimental results in Table 12 indicate that the logistic regression model outperformed six non-linear models in classifying NAFLD versus non-NAFLD, including AdaBoost, GBDT, Naïve Bayes, Neural Network, Random Forest, and SVM.

4. Discussion

NAFLD has gradually become a serious threat to human health owing to its increasing global incidence and wide distribution at different ages. Although some studies have found NAFLD to be in association with the onset of metabolic diseases such as obesity and diabetes [49], other studies have shown that 40.8% of global fatty liver patients are non-obese people, and even 19.2% are “skinny” [50], whereas in Asian populations, approximately 8%–19% of NAFLD patients have a normal BMI [51], suggesting that obesity is not the only criterion for NAFLD screening. In this study, basic indicators such as BMI, waist circumference, sex, and age were selected as the diagnostic elements of NAFLD for analysis. The accuracy of the model for identifying NAFLD was found to be 75.2% (95% CI of 0.819–0.853). Among the NAFLD population included in this study, 27.71% had a normal BMI, which was consistent with the results reported in the literature. The addition of objective tongue image parameters increased the accuracy of model 1to 76.4% (95% CI of 0.825–0.860). Although the accuracy rate was only improved by 1.2%, there was a statistical difference between the two AUROC results, suggesting that the objective tongue image parameters have a certain degree of contribution to the diagnosis of NAFLD, which is non-invasive, convenient, and fast, and will be more suitable for preliminary large-scale NAFLD screening.

After the addition of the serological indicators, the accuracy of the NAFLD diagnostic model increased from 76.4% to 81.7%, and compared with the other models (including HSI and FLI), an LR of 5.25 was the highest, and an NLR of 0.26 was the lowest, indicating that the model combining basic physiological indexes, serological indicators, and objective tongue image parameters was the best diagnostic model for NAFLD applied in this study. Among the tongue image parameters, TB-L represents the brightness of the tongue, and the higher the brightness L value of the tongue is, the whiter the tongue. TB-b is the color index of the tongue, and the lower the b value is, the bluer the tongue. TC-L represents the brightness of the tongue coating, and the higher the brightness of the tongue coating is, the whiter the tongue coating. The logistic regression results of this study indicate that tongue image parameters TB-b and TC-L are protective factors for NAFLD, and TB-L is a

Table 4
Comparison results of tongue color characteristics between two groups.

	non-NAFLD (n = 947)	NAFLD (n = 831)	*t/Z*	*Cohen's d*	*P*
TB-L	48.36 (46.72–49.75)	48.47 (46.92–50.01)	−0.827	0.039	0.408
TB-a	27.41 (2.47)	27.23 (2.46)	1.550	0.073	0.121
TB-b	9.38 (8.29–10.9)	8.69 (7.45–9.9)	−8.552	0.414	<0.001
TC-L	49.74 (46.35–52.31)	49.57 (46.66–52.38)	−0.465	0.022	0.642
TC-a	19.02 (17.48–20.81)	18.97 (17.27–20.59)	−1.771	0.084	0.077
TC-b	6.40 (5.20–7.77)	5.85 (4.44–7.11)	−6.318	0.303	<0.001
perAll	0.33 (0.26–0.38)	0.35 (0.28–0.41)	−5.386	0.258	<0.001

Table 5
Correlation between tongue parameters and physiological indicators.

Indexes	Sex	Age	BMI	TG	ALT/AST	GGT	HSI	FLI
perAll	0.072[a]	0.076[a]	0.083[a]	0.057	0.010	0.066	0.050	0.100[b]
TB-L	−0.184[b]	−0.059	−0.041	−0.058	−0.125[b]	−0.159[b]	−0.042	−0.098[b]
TB-a	0.097[b]	−0.166[b]	0.010	0.031	0.115[b]	0.044	−0.031	0.046
TB-b	−0.286[b]	0.108[b]	−0.182[b]	−0.178[b]	−0.202[b]	−0.203[b]	−0.093[b]	−0.282[b]
TC-L	−0.120[b]	0.048	−0.003	−0.001	−0.071[a]	−0.050	0.044	−0.019
TC-a	0.066	−0.128[b]	0.016	−0.018	0.099[b]	0.043	0.035	0.026
TC-b	−0.182[b]	0.178[b]	−0.136[b]	−0.136[b]	−0.166[b]	−0.106[b]	−0.062	−0.197[b]
Spotted	−0.074[a]	−0.134[b]	−0.035	−0.009	0.041	0.002	−0.030	−0.047

[a] When the confidence level (double tails) is 0.05, the correlation is significant.
[b] When the confidence level (double tails) is 0.01, the correlation is significant.

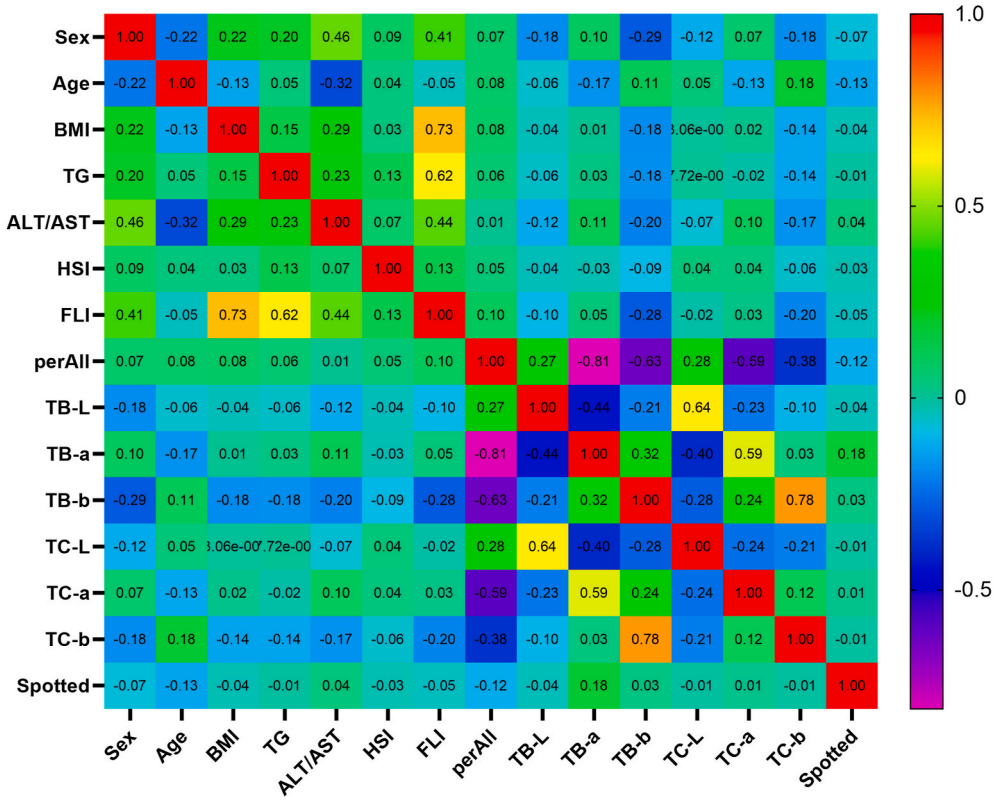

Fig. 6. Correlation between tongue image parameters and physiological indexes.

risk factor. Meanwhile, the TB-b value in the NAFLD group was significantly lower than that in the non-NAFLD group, suggesting that the objective tongue image parameter TB-b can be used as one of the markers for the diagnosis of NAFLD.

However, as an objective and external manifestation of the state of the human body, the tongue image can reflect the changes in the physiological and pathological state of the human body to a certain extent, but it cannot be used as the main basis for the clinical diagnosis of NAFLD. The results of this study showed that tongue image parameters based on computer imaging technology can improve the accuracy of an NAFLD diagnosis. This is a non-invasive and convenient diagnostic method. Although the accuracy is only improved by a few percentage points when compared with the other models, such improvements will be extremely beneficial in the screening of large samples. Meanwhile, the subjects included in this study were mainly young people, and the ultrasound examination results were mainly mild fatty liver. This is also a possible reason for the small effect size and insignificant increase in the accuracy of our model.

With improvements in detection and diagnostic techniques, some non-invasive testing methods have also been widely used in clinical practice, particularly for the diagnosis of hepatic steatosis, NASH, and fibrosis. The AUROC of AAR [52], FIB-4 [53], BARD core [54], and other diagnostic models based on serological examination for grade 3–4 liver fibrosis were all ≥0.8. Although each model has certain limitations [55], a certain reference value has been found in clinical applications. Therefore, some scholars have proposed that non-invasive methods are sufficient for the diagnosis of NAFLD [56]. The application of an objective TCM tongue diagnosis technology to the construction of NAFLD diagnosis models is an extremely useful exploration. We searched the application of a computer tongue image analysis technology in the auxiliary diagnosis of clinical diseases, and found that the GA_XGBT model constructed using computer tongue image analysis technology and the fusing of tongue image feature parameters achieves an average accuracy rate of 0.81 and an average F1 score of 0.796 in the auxiliary diagnosis of type 2 diabetes [29]. These results are similar to the accuracy and F1 score of the best NAFLD diagnostic model found in this study, which to a certain extent, illustrates the reliability of the study results.

However, the clinical application and diagnostic accuracy of this study need to be verified. In a later stage, the screening of NAFLD based

Table 6
Construction of NAFLD diagnosis model based on logistic regression analysis.

Model	Variables	B	SE	Wald	*P*	OR	95% CI
Model1	Age	−0.009	0.006	2.538	0.111	0.991	0.979–1.002
	Sex	1.284	0.162	62.635	<0.001	3.611	2.628–4.963
	Waist	0.056	0.011	27.605	<0.001	1.058	1.036–1.08
	BMI	0.421	0.034	155.761	<0.001	1.523	1.426–1.627
	Constant	−15.392	0.824	348.693	<0.001		
Model2	Age	−0.008	0.006	1.643	0.200	0.992	0.981–1.004
	Sex	1.494	0.171	76.171	0.000	4.455	3.185–6.231
	Waist	0.054	0.011	25.229	0.000	1.056	1.034–1.079
	BMI	0.409	0.034	145.41	0.000	1.506	1.409–1.609
	TB-b	−0.164	0.036	20.984	0.000	0.849	0.791–0.91
	TC-L	−0.033	0.013	6.300	0.012	0.968	0.944–0.993
	Constant	−11.964	1.155	107.313	0.000		
Model3	Age	−0.004	0.007	0.406	0.524	0.996	0.983–1.009
	Sex	2.195	0.192	130.247	<0.001	8.983	6.162–13.097
	Waist	0.043	0.012	13.165	<0.001	1.044	1.020–1.069
	BMI	0.327	0.037	78.098	<0.001	1.387	1.290–1.492
	TG	0.532	0.089	35.342	<0.001	1.702	1.428–2.029
	GGT	0.028	0.006	24.787	<0.001	1.028	1.017–1.040
	ALT/AST	2.080	0.236	77.908	<0.001	8.007	5.045–12.709
	Constant	−16.369	0.935	306.579	<0.001		
Model4	Age	0.000	0.007	0.002	0.965	1.000	0.987–1.014
	Sex	2.263	0.201	126.789	<0.001	9.611	6.482–14.251
	Waist	0.042	0.012	12.338	<0.001	1.043	1.019–1.068
	BMI	0.323	0.037	75.237	<0.001	1.382	1.284–1.486
	TG	0.532	0.091	34.498	<0.001	1.702	1.425–2.032
	GGT	0.028	0.006	24.547	<0.001	1.028	1.017–1.04
	ALT/AST	2.110	0.238	78.456	<0.001	8.246	5.17–13.151
	TB-L	0.092	0.035	6.788	0.009	1.097	1.023–1.175
	TB-b	−0.105	0.040	7.056	0.008	0.900	0.833–0.973
	TC-L	−0.062	0.019	11.04	0.001	0.940	0.907–0.975
	Constant	−16.902	1.82	86.263	<0.001		

Note: The dependent variable is NAFLD (0 = non-NAFLD, 1 = NAFLD), the control variables included Sex (0 = female, 1 = male), age, and independent variables including BMI, WC, tongue image parameters, and serum index; none of the variables have a collinearity problem (VIF < 5), and use the forward likelihood ratio method. Only the tongue image parameters that can enter the regression equation are included in table ($P < 0.05$).

Table 7
Differences of different NAFLD diagnostic models.

Model	F1-score	AUC	Accuracy	Sensitivity	Specificity	SE	95% CI
Model 1	0.7276	0.837	75.20%	70.88%	78.99%	0.0092	0.819–0.853
Model 2	0.7389	0.842	76.30%	71.84%	80.15%	0.0091	0.825–0.859
Model 3	0.7906	0.895	80.90%	77.02%	84.37%	0.0073	0.879–0.909
Model 4	0.7983	0.897	81.70%	77.62%	85.22%	0.0072	0.882–0.911

Table 8
AUROC comparison of different NAFLD diagnostic models.

Variable	ΔAUC	SE	*z statistic*	*P*	95% CI
Model1:Model2	0.0059	0.0022	2.717	0.0066	0.0016–0.0101
Model3:Model4	0.0027	0.0013	2.103	0.0354	0.0002–0.0053

on model 4 with a cutoff value of 0.6319 and the actual accuracy based on ultrasound and other imaging examinations can better determine the feasibility of the model in the diagnosis and prediction of NAFLD. In a stratified study, we will expand the sample size based on the severity of NAFLD and demonstrate the diagnostic value of model 4. In addition, the image acquisition equipment and analysis methods of an objective tongue diagnosis technology have not been unified, which brings about problems in that different research results cannot be compared, hindering the application of an objective tongue diagnosis technology in disease diagnosis and prediction. The previous experimental basis of this study confirmed that the objective tongue image acquisition equipment, acquisition methods, and data analysis techniques used in this study have high credibility for the diagnosis and prediction of related diseases [29,30,36,57] and are worthy of more in-depth research and verification. Meanwhile, modeling methods have an important impact on the accuracy of disease diagnosis models. Our study found that the accuracy rate obtained using the neural network method was similar to the logistic regression results. Therefore, in future studies, we plan to develop a more advanced deep convolutional neural network model construction method for NAFLD-assisted diagnosis research. In addition, the relationship between the changes in these computer tongue image parameters and the human disease state and mechanism of action have yet to be reported in the literature, which will be explored in the future.

Overall, the application of computer tongue image analysis technology to the screening and diagnosis of NAFLD is useful. The rapid development of digital image technology and mobile technology will provide convenient conditions for the application and popularization of intelligent computer tongue image analysis technology in disease diagnosis, and make it possible to be applied in the large-scale screening of NAFLD and other diseases, and provide a basis for further mechanistic research.

5. Conclusions

The global incidence of NAFLD is high and continues to increase, and thus it is extremely important to find a convenient, economical, and accurate NAFLD screening method. The development of computer imaging technology provides convenient conditions for the application of an objective tongue diagnosis.

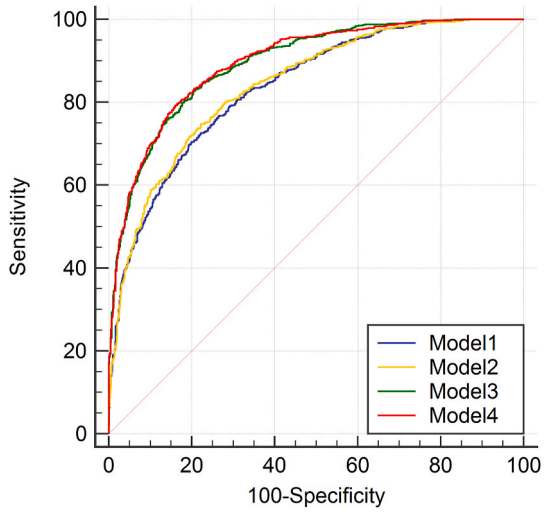

Fig. 7. ROC diagrams of different NAFLD diagnostic models. Model 1 contains sex, age, waist circumference, and BMI; Model 2 contains sex, age, waist circumference, BMI,TB-b, and TC-L; Model 3 contains sex, age, waist circumference, BMI, TG, GGT, and ALT/AST; Model 4 contains sex, age, waist circumference, BMI, TG, GGT, ALT/AST, TB-b, TB-L, and TC-L.

In this study, we found the best diagnostic model suitable for large-scale NAFLD screening by constructing different NAFLD diagnostic models, and confirmed the application value of intelligent tongue image parameters in the diagnosis of NAFLD, providing strong support for the follow-up application of computer imaging technology in disease diagnosis and objective TCM diagnosis technology.

The NAFLD diagnosis model combined with computer tongue image analysis technology is worthy of further exploration.

Consent for publication

All the authors have read and approved the manuscript in all respects for publication.

Funding

This study was supported by the National Key Technology R&D Program of China [grant numbers 2017YFC1703301], the National

Table 11
Comparison of FLI and HSI and model 4 on NAFLD classification performance.

Variable	ΔAUC	SE	*z statistic*	*P*	95% CI
FLI: HSI	0.0057	0.0056	1.014	0.3108	−0.0053 to 0.0166
HSI: Model4	0.0209	0.0038	5.517	<0.001	0.0135 to 0.0283
FLI: Model4	0.0152	0.0033	4.563	<0.001	0.0087 to 0.0218

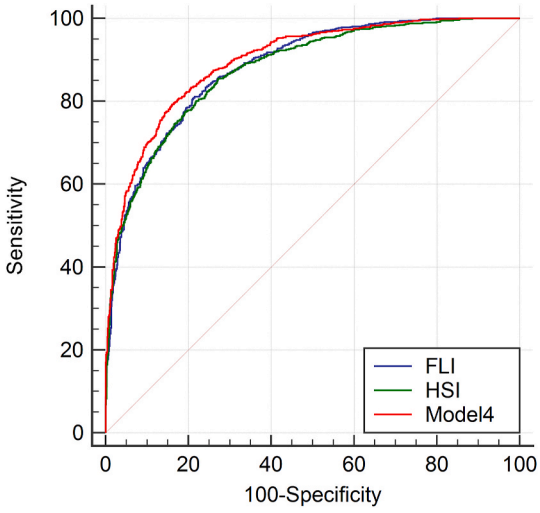

Fig. 8. Results of AUROC comparison for model 4 and FLI and HSI diagnostic models.

Table 12
Comparison of diagnostic efficiency of different NAFLD modeling methods.

Model	F1-score	AUC	Accuracy	Sensitivity	Specificity
AdaBoost	0.7035	0.721	72.22%	70.52%	73.71%
GBDT	0.7762	0.885	79.47%	76.17%	82.37%
Naïve Bayes	0.7490	0.847	76.32%	75.57%	76.98%
Neural Network	0.7886	0.896	80.76%	76.77%	84.27%
Random Forest	0.7574	0.859	77.56%	74.97%	79.83%
SVM	0.7774	0.867	78.07%	81.95%	74.66%

Table 9
Construction of NAFLD diagnostic model on the basis of FLI and HSI.

Model	Variables	B	S.E.	Wald	*P*	OR	95% CI
FLI	Age	−0.021	0.006	11.273	0.001	0.979	0.967–0.991
	Sex	1.883	0.182	107.352	<0.001	6.574	4.604–9.386
	Waist	0.048	0.011	18.076	<0.001	1.05	1.026–1.073
	BMI	0.355	0.036	98.788	<0.001	1.427	1.330–1.530
	TG	0.640	0.089	51.964	<0.001	1.896	1.593–2.256
	GGT	0.046	0.005	73.356	<0.001	1.047	1.036–1.058
	Constant	−15.100	0.879	294.953	<0.001		
HSI	Age	0.009	0.006	2.128	0.145	1.009	0.997–1.022
	BMI	0.453	0.029	245.468	<0.001	1.574	1.487–1.665
	ALT/AST	2.826	0.216	171.52	<0.001	16.886	11.062–25.777
	Sex	1.704	0.168	103.261	<0.001	5.498	3.958–7.638
	Diabetes	0.854	0.300	8.118	0.004	2.349	1.305–4.227
	Constant	−15.433	0.813	360.101	<0.001		

Table 10
The AUROC comparison between FLI and HSI.

Model	F1-s core	AUC	Accuracy	Sensitivity	Specificity	SE	95% CI
FLI	0.7655	0.882	78.70%	74.25%	82.68%	0.0077	0.866–0.897
HSI	0.7696	0.877	79.10%	75.33%	82.47%	0.0080	0.860–0.891

Natural Science Foundation of China [grant numbers 81574058, 81904094 and 81973750], Shanghai Municipal Commission of Health and Family Planning [grant numbers ZY (2018–2020)-CCCX-2001-01] and Shanghai Municipal Health Commission [grant numbers 201940117 and 2020JQ003].

Ethics approval

This study has been approved by the ethics committee of Shuguang Hospital Affiliated to Shanghai University of TCM. The ethics committee number is 2018-626-55-01, and the clinical trial registration number is ChiCTR1900026008.

Declaration of competing interest

None of the authors have conflicts of interest for the manuscript of "Application of computer tongue image analysis technology in the diagnosis of NAFLD".

Acknowledgement

Tao Jiang and Xiao-jing Guo contributed equally to this work, and participated in experimental research design, case collection, data processing and paper writing, Zhou Lu, Xu-xiang Ma, Xing-hua Yao participated in the case collection, Li-Ping Tu, Ji Cui, Xiao-juan Hu and Long-tao Cui participated in data collation and methodological guidance, Yong Zhi Li participated in research design guidance and paper revision, Jia-tuo Xu and Jing-bin Huang participated in the design and implementation of the research, and provided important guidance for data processing, writing and revision of the paper.

Appendices.

The formulas used in the research are as follows:

$$\text{Accuracy: Acc} = \frac{TP + TN}{TN + TP + FN + FP} \tag{A.1}$$

$$\text{Sensitivity/Recall: } Sens \Big/ R = \frac{TP}{TP + FN} \tag{A.2}$$

$$\text{Specificity: } Spec = \frac{TN}{TN + FP} \tag{A.3}$$

$$\text{Positive LR: } LR = \frac{\text{Sens}}{1 - \text{Spec}} \tag{A.4}$$

$$\text{Negative LR: } NLR = \frac{1 - \text{Sens}}{\text{Spec}} \tag{A.5}$$

$$\text{Precision: } P = \frac{TP}{TP + FP} \tag{A.6}$$

$$\text{F1} - \text{Score: } F1 = \frac{2{*}\text{P} * \text{R}}{\text{P} + \text{R}} \tag{A.7}$$

$$\text{FLI} = \frac{e^{0.953*\log e\ (\text{triglycerides}) + 0.139*\text{BMI} + 0.718*\log e\ (\text{GGT}) + 0.053*\text{waist circumference} - 15.745}}{1 + e^{0.953*\log e\ (\text{triglycerides}) + 0.139*\text{BMI} + 0.718*\log e\ (\text{GGT}) + 0.053*\text{waist circumference} - 15.745}} \times 100$$

$$\text{HSI} = 8 \times \frac{\text{ALT}}{\text{AST}} + \text{BMI}(+2,\ \text{if DM};\ +2,\ \text{if female})$$

References

[1] M.S. Mundi, S. Velapati, J. Patel, et al., Evolution of NAFLD and its management, Nutr. Clin. Pract. 35 (1) (2020) 72–84, https://doi.org/10.1002/ncp.10449.
[2] Z.M. Younossi, A.B. Koenig, D. Abdelatif, et al., Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes, Hepatology 64 (1) (2016) 73–84, https://doi.org/10.1002/hep.28431.
[3] C.C. Lindenmeyer, A.J. McCullough, The natural history of nonalcoholic fatty liver disease-an evolving view, Clin. Liver Dis. 22 (1) (2018) 11–21, https://doi.org/10.1016/j.cld.2017.08.003.
[4] R. Lombardi, S. Onali, D. Thorburn, et al., Pharmacological interventions for non-alcohol related fatty liver disease (NAFLD): an attempted network meta-analysis, Cochrane Database Syst. Rev. 3 (3) (2017) Cd011640, https://doi.org/10.1002/14651858.CD011640.pub2.
[5] EASL-EASD-EASO Clinical Practice Guidelines for the management of non-alcoholic fatty liver disease, J. Hepatol. 64 (6) (2016) 1388–1402, https://doi.org/10.1016/j.jhep.2015.11.004.
[6] B. Macavei, A. Baban, D.L. Dumitrascu, Psychological factors associated with NAFLD/NASH: a systematic review, Eur. Rev. Med. Pharmacol. Sci. 20 (24) (2016) 5081–5097.
[7] L. Pacifico, F.M. Perla, M. Roggini, et al., A systematic review of NAFLD-associated extrahepatic disorders in youths, J. Clin. Med. 8 (6) (2019), https://doi.org/10.3390/jcm8060868.
[8] T.G. Simon, B. Roelstraete, H. Khalili, et al., Mortality in biopsy-confirmed nonalcoholic fatty liver disease: results from a nationwide cohort, Gut (2020), https://doi.org/10.1136/gutjnl-2020-322786.
[9] D.A. Koutoukidis, N.M. Astbury, K.E. Tudor, et al., Association of weight loss interventions with changes in biomarkers of nonalcoholic fatty liver disease: a systematic review and meta-analysis, JAMA internal medicine 179 (9) (2019) 1262–1271, https://doi.org/10.1001/jamainternmed.2019.2248.
[10] E. Vilar-Gomez, Y. Martinez-Perez, L. Calzadilla-Bertot, et al., Weight loss through lifestyle modification significantly reduces features of nonalcoholic steatohepatitis, Gastroenterology 149 (2) (2015) 367–378, https://doi.org/10.1053/j.gastro.2015.04.005, e5; quiz e14-5.
[11] G. Bedogni, S. Bellentani, L. Miglioli, et al., The Fatty Liver Index: a simple and accurate predictor of hepatic steatosis in the general population, BMC Gastroenterol. 6 (2006) 33, https://doi.org/10.1186/1471-230x-6-33.
[12] J.H. Lee, D. Kim, H.J. Kim, et al., Hepatic steatosis index: a simple screening tool reflecting nonalcoholic fatty liver disease, Dig. Liver Dis. 42 (7) (2010) 503–508, https://doi.org/10.1016/j.dld.2009.08.002.
[13] L.A. Rabbitt, M. McNally, L. Reynolds, et al., A prospective cohort study of the use of the fatty liver index and Fibroscan to determine the prevalence of fatty liver

disease in an Irish population, Eur. J. Gastroenterol. Hepatol. (2020), https://doi.org/10.1097/meg.0000000000001951.
[14] L. Sviklāne, E. Olmane, Z. Dzērve, et al., Fatty liver index and hepatic steatosis index for prediction of non-alcoholic fatty liver disease in type 1 diabetes, J. Gastroenterol. Hepatol. 33 (1) (2018) 270–276, https://doi.org/10.1111/jgh.13814.
[15] P.Y. Liao, P.C. Hsu, J.M. Chen, et al., Diabetes with pyogenic liver abscess–A perspective on tongue assessment in traditional Chinese medicine, Compl. Ther. Med. 22 (2) (2014) 341–348, https://doi.org/10.1016/j.ctim.2013.12.009.
[16] Y. Chen, T.H. Jiang, W.Z. Ru, et al., Objective tongue inspection on 142 liver cancer patients with damp-heat syndrome, Chin. J. Integr. Med. 20 (8) (2014) 585–590, https://doi.org/10.1007/s11655-014-1756-z.
[17] L.C. Lo, C.Y. Chen, J.Y. Chiang, et al., Tongue diagnosis of traditional Chinese medicine for rheumatoid arthritis, Afr. J. Tradit., Complementary Altern. Med. : AJTCAM 10 (5) (2013) 360–369, https://doi.org/10.4314/ajtcam.v10i5.24.
[18] L.C. Lo, T.L. Cheng, J.Y. Chiang, et al., Breast cancer index: a perspective on tongue diagnosis in traditional Chinese medicine, Journal of traditional and complementary medicine 3 (3) (2013) 194–203, https://doi.org/10.4103/2225-4110.114901.
[19] T.C. Wu, C.N. Lu, W.L. Hu, et al., Tongue diagnosis indices for gastroesophageal reflux disease: a cross-sectional, case-controlled observational study, Medicine (Baltim.) 99 (29) (2020), e20471, https://doi.org/10.1097/md.0000000000020471.
[20] Z. Qi, L.P. Tu, J.B. Chen, et al., The classification of tongue colors with standardized acquisition and ICC profile correction in traditional Chinese medicine, BioMed Res. Int. 2016 (2016) 3510807, https://doi.org/10.1155/2016/3510807.
[21] M.H. Tania, K. Lwin, M.A. Hossain, Advances in automated tongue diagnosis techniques, Integr Med Res 8 (1) (2019) 42–56, https://doi.org/10.1016/j.imr.2018.03.001.
[22] C.J. Jung, Y.J. Jeon, J.Y. Kim, et al., Review on the current trends in tongue diagnosis systems, Integrative medicine research 1 (1) (2012) 13–20, https://doi.org/10.1016/j.imr.2012.09.001.
[23] M.J. Shi, G.Z. Li, F.F. Li, et al., Computerized tongue image segmentation via the double geo-vector flow, Chin. Med. 9 (1) (2014) 7, https://doi.org/10.1186/1749-8546-9-7.
[24] S. Naveed, G. Geetha, Intelligent diabetes detection system based on tongue datasets, Curr. Med. Imag. Rev. 15 (7) (2019) 672–678, https://doi.org/10.2174/1573405614666181009133414.
[25] M.C. Hu, M.H. Cheng, K.C. Lan, Color correction parameter estimation on the smartphone and its application to automatic tongue diagnosis, J. Med. Syst. 40 (1) (2016) 18, https://doi.org/10.1007/s10916-015-0387-z.
[26] J.M. Lee, S.Y. Yoo, Y.B. Park, Reliability and validity of tongue color analysis in the prediction of symptom patterns in terms of East Asian Medicine, J. Tradit. Chin. Med. 36 (2) (2016) 165–172, https://doi.org/10.1016/s0254-6272(16)30023-1.
[27] M. Zhu, J. Du, C. Ding, A comparative study of contemporary color tongue image extraction methods based on HSI, Int. J. Biomed. Imag. 2014 (2014) 534507, https://doi.org/10.1155/2014/534507.
[28] P.C. Hsu, H.K. Wu, Y.C. Huang, et al., The tongue features associated with type 2 diabetes mellitus, Medicine (Baltim.) 98 (19) (2019), e15567, https://doi.org/10.1097/md.0000000000015567.
[29] J. Li, P. Yuan, X. Hu, et al., A tongue features fusion approach to predicting prediabetes and diabetes with machine learning, J. Biomed. Inf. 115 (2021) 103693, https://doi.org/10.1016/j.jbi.2021.103693.
[30] L. Jun, C. Qingguang, H. Xiaojuan, et al., Establishment of noninvasive diabetes risk prediction model based on tongue features and machine learning techniques, Int. J. Med. Inf. (2021) 104429, https://doi.org/10.1016/j.ijmedinf.2021.104429.
[31] L.C. Lo, T.L. Cheng, Y.J. Chen, et al., TCM tongue diagnosis index of early-stage breast cancer, Compl. Ther. Med. 23 (5) (2015) 705–713, https://doi.org/10.1016/j.ctim.2015.07.001.
[32] T.C. Lee, L.C. Lo, F.C. Wu, Traditional Chinese Medicine for Metabolic Syndrome via TCM Pattern Differentiation: Tongue Diagnosis for Predictor. Evidence-Based Complementary and Alternative Medicine : eCAM, vol. 2016, 2016, p. 1971295, https://doi.org/10.1155/2016/1971295.
[33] Guidelines of prevention and treatment for nonalcoholic fatty liver disease:a 2018 update, J Prac Hepatol 21 (2) (2018) 177–186.
[34] J.F. Junping Shi, Expert recommendations on standardized diagnosis and treatment for fatty liver disease (2019 revised edition), J Prac Hepatol 22 (6) (2019) 787–792.
[35] W. Jiao, X.J. Hu, L.P. Tu, et al., Tongue color clustering and visual application based on 2D information, Int J Comput Assist Radiol Surg 15 (2) (2020) 203–212, https://doi.org/10.1007/s11548-019-02076-z.
[36] J. Zhang, J. Xu, X. Hu, et al., Diagnostic method of diabetes based on support vector machine and tongue images, BioMed Res. Int. 2017 (2017) 7961494, https://doi.org/10.1155/2017/7961494.
[37] L. Zheng, X. Zhang, J. Hu, et al., Establishment and applicability of a diagnostic system for advanced gastric cancer T staging based on a faster region-based convolutional neural network, Frontiers in oncology 10 (2020) 1238, https://doi.org/10.3389/fonc.2020.01238.
[38] S. Ren, K. He, R. Girshick, et al., Faster R-CNN: towards real-time object detection with region proposal networks, IEEE Trans. Pattern Anal. Mach. Intell. 39 (6) (2017) 1137–1149, https://doi.org/10.1109/tpami.2016.2577031.
[39] R. Li, X. Zeng, S.E. Sigmund, et al., Automatic localization and identification of mitochondria in cellular electron cryo-tomography using faster-RCNN, BMC Bioinf. 20 (Suppl 3) (2019) 132, https://doi.org/10.1186/s12859-019-2650-7.
[40] T. Jiang, X.J. Hu, X.H. Yao, et al., Tongue image quality assessment based on a deep convolutional neural network, BMC Med. Inf. Decis. Making 21 (1) (2021) 147, https://doi.org/10.1186/s12911-021-01508-8.
[41] X. Jiatuo, T. Liping, Z. Zhifeng, et al., Identification of tongue body and Fur based on color image region separation, Journal of Shanghai University of traditional Chinese Medicine 23 (3) (2009) 42–45.
[42] P. Jackman, D.W. Sun, G. Elmasry, Robust colour calibration of an imaging system using a colour space transform and advanced regression modelling, Meat Sci. 91 (4) (2012) 402–407, https://doi.org/10.1016/j.meatsci.2012.02.014.
[43] W. Lenhard, A. Lenhard, Calculation of effect sizes (Germany): Psychometrica. DOI: 10.13140/RG.2.2.17823.92329, https://www.psychometrica.de/effect_size.html. Dettelbach, 2016.
[44] Q. Ren, X.W. Zhou, M.Y. He, et al., A quantitative diagnostic method for phlegm and blood stasis syndrome in coronary heart disease using tongue, face, and pulse indexes: an exploratory pilot study, J. Alternative Compl. Med. 26 (8) (2020) 729–737, https://doi.org/10.1089/acm.2020.0008.
[45] Q.Q. Wang, S.C. Yu, X. Qi, et al., Overview of logistic regression model analysis and application, Zhonghua yu fang yi xue za zhi [Chinese journal of preventive medicine] 53 (9) (2019) 955–960, https://doi.org/10.3760/cma.j.issn.0253-9624.2019.09.018.
[46] R. Raghavan, F.S. Ashour, R. Bailey, A review of cutoffs for nutritional biomarkers, Advances in nutrition (Bethesda, Md) 7 (1) (2016) 112–120, https://doi.org/10.3945/an.115.009951.
[47] D.M.W. Powers, Evaluation: from precision, recall and fmeasure to roc, informedness, markedness and correlation, J. Mach. Learn. Technol. 2 (1) (2011) 37–63.
[48] A. Hirata, D. Sugiyama, K. Kuwabara, et al., Fatty liver index predicts incident diabetes in a Japanese general population with and without impaired fasting glucose, Hepatol. Res. 48 (9) (2018) 708–716, https://doi.org/10.1111/hepr.13065.
[49] B.J. Perumpail, M.A. Khan, E.R. Yoo, et al., Clinical epidemiology and disease burden of nonalcoholic fatty liver disease, World J. Gastroenterol. 23 (47) (2017) 8263–8276, https://doi.org/10.3748/wjg.v23.i47.8263.
[50] Q. Ye, B. Zou, Y.H. Yeo, et al., Global prevalence, incidence, and outcomes of non-obese or lean non-alcoholic fatty liver disease: a systematic review and meta-analysis, Lancet Gastroenterol Hepatol 5 (8) (2020) 739–752, https://doi.org/10.1016/s2468-1253(20)30077-7.
[51] J.G. Fan, S.U. Kim, V.W. Wong, New trends on obesity and NAFLD in Asia, J. Hepatol. 67 (4) (2017) 862–873, https://doi.org/10.1016/j.jhep.2017.06.003.
[52] A. Kotronen, M. Peltonen, A. Hakkarainen, et al., Prediction of non-alcoholic fatty liver disease and liver fat using metabolic and genetic factors, Gastroenterology 137 (3) (2009) 865–872, https://doi.org/10.1053/j.gastro.2009.06.005.
[53] P. Angulo, J.M. Hui, G. Marchesini, et al., The NAFLD fibrosis score: a noninvasive system that identifies liver fibrosis in patients with NAFLD, Hepatology 45 (4) (2007) 846–854, https://doi.org/10.1002/hep.21496.
[54] J.K. Dyson, Q.M. Anstee, S. McPherson, Non-alcoholic fatty liver disease: a practical approach to diagnosis and staging, Frontline Gastroenterol. 5 (3) (2014) 211–218, https://doi.org/10.1136/flgastro-2013-100403.
[55] J. Aron-Wisnewsky, C. Vigliotti, J. Witjes, et al., Gut microbiota and human NAFLD: disentangling microbial signatures from metabolic disorders, Nat. Rev. Gastroenterol. Hepatol. (2020), https://doi.org/10.1038/s41575-020-0269-9.
[56] L. Castera, Diagnosis of non-alcoholic fatty liver disease/non-alcoholic steatohepatitis: non-invasive tests are enough, Liver Int. 38 (Suppl 1) (2018) 67–70, https://doi.org/10.1111/liv.13658.
[57] Z. Qi, L.P. Tu, Z.Y. Luo, et al., Tongue Image Database Construction Based on the Expert Opinions: Assessment for Individual Agreement and Methods for Expert Selection. Evidence-Based Complementary and Alternative Medicine : eCAM, vol. 2018, 2018, p. 8491057, https://doi.org/10.1155/2018/8491057.