A Study On Fatigue Classification By Logistic Regression Method: Based On Data of Tongue And Pulse

Yulin Shi (yilinlife94@126.com)
Shanghai University of Traditional Chinese Medicine https://orcid.org/0000-0003-4379-8561

Tao Jiang
Shanghai University of Traditional Chinese Medicine

Xiao-juan Hu
Shanghai University of Traditional Chinese Medicine

Ji Cui
Shanghai University of Traditional Chinese Medicine

Long-tao Cui
Shanghai University of Traditional Chinese Medicine

Li-ping Tu
Shanghai University of Traditional Chinese Medicine

Xing-hua Yao
Shanghai University of Traditional Chinese Medicine

Fang Lan
Shanghai University of Traditional Chinese Medicine

Jun Li
Shanghai University of Traditional Chinese Medicine

Zi-jaun Bi
Shanghai University of Traditional Chinese Medicine

Jia-cai Li
Shanghai University of Traditional Chinese Medicine

Jing-bin Huang
Shanghai University of Traditional Chinese Medicine

Jia-tuo Xu
Shanghai University of Traditional Chinese Medicine

Research Article

Keywords: fatigue, tongue diagnosis, pulse diagnosis, logistic regression analysis, intelligent diagnosis

Posted Date: June 11th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-551999/v1

Abstract

Purpose: Fatigue is a subjective symptom which is hard to quantify, it is prevalent in sub-health and disease population, and there is still no accurate and stable method to distinguish disease fatigue from sub-health fatigue. Tongue and pulse diagnoses are the reflection of the overall state of the body, and the modern researches of tongue and pulse diagnoses have made great progress. This study aims to explore the distribution rules of tongue and pulse data in disease fatigue and sub-health fatigue population, and evaluate the contribution rate of tongue and pulse data to fatigue diagnosis through modeling.

Methods: Tongue and Face Diagnosis Analysis-1 instrument and Pulse Diagnosis Analysis-1 instrument were used to collect tongue image and pulse sphygmogram of sub-health fatigue population (n=252) and disease fatigue group(n=1160), we mainly analyzed the tongue and pulse characteristics and constructed the classification model based on logistic regression method.

Results: The results showed that sub-health fatigue people and disease fatigue people had different characteristics of tongue and pulse, and the logistic regression model based on tongue and pulse data showed a better classification effect. The accuracy of healthy controls & sub-health fatigue, sub-health fatigue & disease fatigue, health controls & disease fatigue model were 65.70%, 65.10%, 78.90%, and the AUC were 0.678, 0.834, and 0.879 respectively.

Conclusion: This study provided a new non-invasive method for the fatigue diagnosis from the perspective of objective tongue and pulse data, the modern tongue and pulse diagnoses have a good application prospect.

Introduction

Fatigue refers to symptom that physical tiredness with lack of energy or mental exhaustion with lack of concentration. It can be divided into physical fatigue and mental fatigue[20]. Fatigue is the first cause of sub-health, and is one of the most common symptoms in primary care, and it is experienced by many patients with chronic hepatitis[6, 22], depression[3], and various types of cancers[4]. Sub-health and a wide variety of diseases are associated with different degrees of fatigue with negative effect on people's life. With the improvement of general medical care and living standard, fatigue is more and more awared by people, however, due to the lack of objective diagnostic evidence of fatigue, there is still no reliable and stable evaluation method to distinguish disease fatigue and sub-health fatigue.

A large number of clinical practices and studies have shown that tongue and pulse can reflect the overall state of body[24]. Intelligent diagnosis of TCM is a new research field in recent years, and it meets the trend that TCM diagnosis methods developing gradually towards intelligence and potential application in clinical practice[19, 10]. In recent researches of tongue and pulse diagnoses, new diagnosis systems are adopted to collect and analyze clinical data related to disease, and machine learning method such as Artificial Neural Network[16, 23], Support Vector Machine(SVM) [26, 9] and KNN[25] are used to establish corresponding diagnosis model, which can effectively assist doctor on the diagnosis of disease. There are more and more studies on fatigue based on tongue and pulse diagnoses[15, 21, 17].

Based on the modern research of tongue and pulse diagnoses, this study aims to explore the distribution rules of the data of tongue and pulse in disease fatigue and sub-health fatigue, and evaluate the contribution rate of the data to fatigue diagnosis through modeling, so as to provide a new reference for convenient and non-invasive methods of

fatigue diagnosis, if an objective evaluation method based on the data of tongue and pulse can be established, it will play an important role in clinical diagnosis of fatigue.

Methods

2.1 Study design

A total of 7,025 subjects were collected from January 2015 to December 2018 in the medical examination center of Shuguang Hospital Affiliated to Shanghai University of Traditional Chinese Medicine, collecting their Western medicine physical examination index, tongue and pulse data of TCM. The 7,025 subjects were divided into healthy controls (n = 799), sub-health fatigue group (n = 361) and disease fatigue group (n = 1529). After excluding the outliers with extreme values in tongue or pulse data, there were 551, 252, and 1,160 subjects in healthy controls, sub-health fatigue group and disease fatigue group respectively. The overall flow diagram of the study was shown as Fig. 1.

2.2 Diagnostic criteria

Health and sub-health state of each individual were determined using the Health Status Assessment Questionnaire H20 Scale[8] and the Information Record Form of Four Diagnosis of TCM[12] (Copyright No. : 2016Z11L025702) which were designed by the Sub-Health Research Group. The diagnostic criteria of disease were showed as Table 1.

Table 1
Diagnostic criteria of disease

Disease	Diagnostic Criteria
Diabetes[7]	Fasting blood glucose ≥ 7.0mmol/L and/or blood glucose at any point ≥ 7.8mmol/L and/or blood glucose at two hours after meal ≥ 11.1mmol/L
Hypertension[18]	Systolic blood pressure ≥ 140 mmHg and/or Diastolic blood pressure ≥ 90 mmHg
Hyperlipidemia[2]	TC ≥ 6.2mmol/L and/or LDL-C ≥ 4.1mmol/L and/or HDL-C ≥ 4.9mmol/L and/or TG ≥ 2.3mmol/L and/or non-HDL-C ≥ 1.55mmol/L
Fatty liver disease[5]	Ultrasound examination

Disease was diagnosed by four well-trained clinicians according to the above diagnostic criteria of Western medicine. After excluding the disease population, the population with a score between 60 and 79 on the H20 scale was sub-health population, and a score between 80 and 100 on the H20 scale was healthy controls. Finally, the Information Record Form of Four Diagnosis of TCM and H20 scale were used to select fatigue population.

2.3 Tongue diagnosis and Pulse diagnosis Instruments

TFDA-1 tongue and face diagnosis instrument[13] and PDA-1 pulse diagnosis instrument[11] were shown in Fig. 2 and Fig. 3, they were used for data collection. The indexes of tongue image from color spaces of RGB, HSI, Lab and YCrCb. The prefix TB represented the tongue body index, TC represented the tongue coating index. Each of the index of pulse has its meaning[19].

2.4 Data Analysis

SPSS (Version 23.0) software was used for statistical analysis of the data. The normal distribution measurement data were expressed as "X ± S". Non-normal distribution data are expressed as quartiles expressed as "median (upper

quartile, lower quartile)". Analysis of Variance (ANOVA) was performed for normality and homogeneity of variance among groups, Kruskal-wallis H test was performed for non-normal distribution data, and GraphPad Prism Version 8.0 was used for violin plot. All the test result was double-tailed test, test level was $\alpha = 0.05$, and the difference was statistically significant when $P < 0.05$.

2.5 Modeling

Logistic regression analysis is performed for factors with statistical significance by ANOVA or Rank Sum Test. It is often used in data mining, automatic disease diagnosis, economic prediction and others, accuracy of the decision can be improved by adjusting the parameters of the regression model[1, 27]. Accuracy, Sensitivity and Specificity are used to evaluate the performance of models. Area under the receiver operator characteristic (ROC) curve was also used to evaluate models, which generally has value between 0.5-1, the larger the value, the better the effect of the classification. Accuracy is the most common evaluation index, which is the ratio of the number of samples correctly classified by the model to the total number of samples. The higher the index, the better the performance of the classifier. Sensitivity is the true positive rate, that is, the percentage of people with actual disease who are correctly diagnosed. Specificity also known as true negative rate, it reflects the ability of a test to identify non-patients. Accuracy, Sensitivity, and Specificity were defined as follows:

$$\text{Accuracy} = \frac{\text{TP}+\text{TN}}{TP+TN+FP+FN} \times 100\% \quad (1)$$

$$\text{Sensitivity} = \frac{\text{TP}}{TP+FN} \times 100\% \quad (2)$$

$$\text{Specificity} = \frac{\text{TN}}{TN+FP} \times 100\% \quad (3)$$

In the above formulas, TP represents the true positive rate, TN represents the true negative rate, FP represents the false positive rate, FN represents the false negative rate.

Results

3.1 General Result

Diseases in disease fatigue group mainly included hypertension, diabetes, hyperlipidemia and fatty liver with distribution were shown in Fig. 4. Table 2 showed the general result of the healthy controls, the group of sub-health fatigue and disease fatigue.

Table 2
General result

Group		N	Male	female	Age	BMI
			N(%)	N(%)	(x ± s, year)	(Kg/▯)
Healthy controls		551	394(71.5)	157(28.5)	29.00(25.00,35.00)	22.31(20.55,24.51)
Sub-health fatigue		252	149(59.1)	103(40.9)	32.00(28.00,37.00)*	22.39(20.28,24.68)
Disease	Hypertension	219	160(73.1)	59(26.9)	46.00(36.00,57.00)**##	25.30(23.40,27.40)**##
	Diabetes	103	84(81.6)	19(18.4)	56.00(45.00,63.00)**##	25.20(23.50,27.70)**##
	Hyperlipemia	388	272(70.1)	116(29.9)	43.50(34.00,53.00)**##	24.90(22.90,26.90)**##
fatigue	Fatty liver	442	334(75.6)	108(24.4)	45.00(34.00,55.00)**##	26.20(24.60,28.10)**##
vs. Healthy controls, *P < 0.05, vs. Healthy controls, ** P < 0.01.						

vs. Sub-health fatigue, #P < 0.05, vs. Sub-health fatigue, ##P < 0.01.

3.2 Statistical Analysis of Tongue indexes

Table 3 showed the statistical analysis result of the distribution of the characteristic parameters of tongue body and tongue coating among the healthy controls, the group of sub-health fatigue and disease fatigue.

Table 3
Statistical result of tongue body and tongue coating among the three groups

Domain	Color space	Index	Healthy controls	Sub-health fatigue	Disease fatigue
			(N = 551)	(N = 252)	(N = 1160)
TB	RGB	TB-R	156.00(143.00,171.00)	153.00(139.00,165.75)	158.00(147.00,169.00)$^{\#\#}$
		TB-G	99.00(90.00,110.00)	98.00(87.00,110.00)	100.00(92.00,110.75)$^{\#}$
		TB-B	105.00(94.00,118.00)	104.00(93.00,118.00)	108.00(99.00,120.00)$^{**\#\#}$
	YCrCb	TB-Y	115.77(107.05,126.27)	115.40(105.14,125.09)	117.46(110.08,126.16)$^{\#\#}$
		TB-Cr	152.17(149.64,154.65)	151.31(148.37,153.84)**	152.21(148.82,155.17)$^{\#}$
		TB-Cb	121.46(119.99,124.21)	121.32(119.84,126.08)	121.74(120.00,127.46)*
	HSI	TB-H	176.27(169.98,179.15)	176.52(166.25,180.00)	175.16(162.12,178.32)$^{**\#\#}$
		TB-S	0.18(0.16,0.20)	0.18(0.15,0.20)	0.18(0.16,0.20)
		TB-I	119.00(108.00,132.00)	118.50(106.00,130.00)	121.00(113.00,132.00)$^{*\#\#}$
	Lab	TB-L	104.33(100.72,108.58)	104.08(99.70,108.05)	104.94(101.84,108.42)$^{\#\#}$
		TB-a	20.62(18.42,22.79)	20.49(18.11,22.68)	20.98(18.69,23.13)
		TB-b	4.91(1.64,6.59)	5.01(-0.17,6.87)	4.35(-1.82,6.23)$^{**\#}$
	texture index	TB-CON	65.33(42.36,92.43)	66.07(40.86,97.57)	60.92(41.26,85.49)
		TB-ENT	1.19(1.09,1.27)	1.19(1.08,1.28)	1.17(1.08,1.26)
		TB-ASM	0.08(0.07,0.10)	0.08(0.07,0.10)	0.08(0.07,0.10)*
		TB-MEAN	0.02(0.02,0.03)	0.02(0.02,0.03)	0.02(0.02,0.03)*
TC	RGB	TC-R	152.00(17.65)	149.56(17.84)	152.44(16.62)
		TC-G	115.00(103.00,126.00)	114.50(103.00,127.00)	114.00(104.00,126.00)
		TC-B	120.00(106.00,134.00)	118.00(105.25,135.00)	121.00(109.00,138.00)*
	YCrCb	TC-Y	124.63(115.36,134.38)	123.65(113.84,134.00)	124.44(116.39,134.07)
		TC-Cr	143.88(141.46,146.59)	143.15(140.51,145.71)*	143.95(140.29,147.04)
		TC-Cb	123.25(121.65,126.17)	123.25(121.64,128.76)	123.76(121.83,129.82)**
	HSI	TC-H	177.27(167.16,181.53)	177.39(156.48,182.40)	175.23(151.57,180.00)$^{**\#\#}$
		TC-S	0.12(0.10,0.14)	0.12(0.09,0.14)	0.12(0.10,0.14)

vs. Healthy controls, $^{*}P < 0.05$, vs. Healthy controls, $^{**}P < 0.01$.

	TC-I	129.00(117.00,140.00)	128.00(115.00,140.00)	129.00(119.00,140.00)
Lab	TC-L	108.41(104.62,112.06)	108.11(104.12,112.09)	108.28(105.06,112.02)
	TC-a	13.10(2.78)	12.71(2.78)	13.27(2.75)
	TC-b	3.17(0.03,4.86)	3.22(-1.97,4.95)	2.63(-3.34,4.68)*
TC index	perAll	0.50(0.41,0.67)	0.53(0.41,0.76)	0.49(0.38,0.79)
	perPart	1.11(1.04,1.25)	1.08(1.03,1.22)	1.10(1.02,1.21)
texture index	TC-CON	82.23(57.65,115.59)	88.97(58.76,125.73)	82.58(56.88,114.81)
	TC-ENT	1.25(1.16,1.32)	1.27(1.17,1.34)	1.25(1.16,1.33)
	TC-ASM	0.07(0.06,0.08)	0.07(0.06,0.08)	0.07(0.06,0.09)
	TC-MEAN	0.03(0.02,0.03)	0.03(0.02,0.03)	0.03(0.02,0.03)

vs. Healthy controls, *P < 0.05, vs. Healthy controls, ** P < 0.01.

vs. Sub-health fatigue, #P < 0.05, vs. Sub-health fatigue group, ##P < 0.01.

Figure 5 showed the Violin Plots of selected parameters of tongue body and tongue coating with statistical significance in the statistical result.

The main result of tongue indexes was: (1) Compared among the three groups, the changes of TB indexes in the group of sub-health fatigue and disease fatigue were more significant than those in the TC indexes. (2) Indexes difference was more significant between the group of disease fatigue and sub-health fatigue. (3) Several indexes (TB-B, TB-R, TB-G, TC-B, TB-I, TB-Y, TB-L, TB-Cb, TB-Cr) in the healthy controls were between the two fatigue groups. It reflected that the two groups of fatigue people had different tendency in the changing nature of tongue.

3.3 Statistical Analysis of Pulse indexes

Table 4 showed the statistical analysis result of the distribution of pulse characteristic parameters in healthy controls, the group of sub-health fatigue and disease fatigue.

Table 4
Statistical result of pulse characteristic parameters of the three groups

Index	Healthy controls	Sub-health fatigue	Disease fatigue
	(N = 551)	(N = 252)	(N = 1160)
t_1(s)	0.13(0.12,0.14)	0.13(0.12,0.14)	0.13(0.12,0.14)$^{**\#\#}$
t_2(s)	0.22(0.21,0.24)	0.22(0.21,0.24)	0.23(0.22,0.24)$^{**\#\#}$
t_3(s)	0.26(0.25,0.27)	0.26(0.25,0.27)	0.26(0.25,0.28)$^{**\#}$
t_4(s)	0.34(0.33,0.36)	0.34(0.33,0.36)	0.35(0.34,0.37)$^{**\#\#}$
t_5(s)	0.40(0.39,0.42)	0.41(0.39,0.42)	0.41(0.39,0.42)
h_1(mv)	112.37(91.47,131.98)	110.54(89.53,132.73)	102.70(79.10,129.20)$^{**\#}$
h_2(mv)	76.26(57.64,96.64)	76.52(60.46,95.92)	72.92(53.31,96.65)
h_3(mv)	68.33(53.88,88.51)	69.70(55.68,87.59)	65.87(48.83,86.66)
h_4(mv)	42.20(32.03,51.17)	41.58(32.52,51.59)	38.25(28.18,49.68)$^{**\#}$
h_5(mv)	3.54(1.08,6.80)	3.35(0.66,6.18)	1.76(0.18,4.22)$^{**\#}$
w_1(s)	0.16(0.13,0.19)	0.17(0.14,0.19)	0.17(0.14,0.20)**
w_2(s)	0.10(0.09,0.13)	0.11(0.09,0.14)	0.12(0.09,0.15)$^{**\#}$
w_1/t	0.20(0.16,0.22)	0.20(0.18,0.23)	0.21(0.18,0.24)$^{**\#}$
w_2/t	0.13(0.11,0.16)	0.14(0.11,0.17)*	0.15(0.12,0.18)$^{**\#}$
h_1/t_1	874.36(701.84,1041.02)	862.66(693.70,1046.20)	775.13(607.61,972.90)$^{**\#\#}$
h_3/h_1	0.62(0.55,0.73)	0.64(0.56,0.73)	0.66(0.56,0.74)**
h_4/h_1	0.38(0.08)	0.38(0.08)	0.37(0.08)
As(mv·s)	0.20(0.03)	0.21(0.03)	0.21(0.03)$^{**\#}$
Ad(mv·s)	0.11(0.09,0.13)	0.11(0.09,0.13)	0.10(0.08,0.12)$^{**\#\#}$
t(s)	0.82(0.75,0.90)	0.82(0.77,0.90)	0.82(0.75,0.90)
vs. Healthy controls, $^{*}P < 0.05$, vs. Healthy controls, ** $P < 0.01$.			

vs. Sub-health fatigue, $^{\#}P < 0.05$, vs. Sub-health fatigue, $^{\#\#}P < 0.01$.

Figure 6 showed the Violin Plots of selected parameters of pulse characteristic with statistical significance.

The main result of pulse feature parameters showed that: t_1, t_2, t_3, t_4, h_1, h_4, h_5, w_1, w_2, w_1/t, w_2/t, h_1/t_1, h_3/h_1, As and Ad had significant statistical differences between the group of disease fatigue and healthy controls ($P < 0.05$, $P < 0.01$), t_4 had significantly statistical differences between the group of sub-health fatigue and the healthy controls (P

< 0.05), t_1, h_1, h_4, h_5, h_1/t_1, Ad, w_1, w_2, w_1/t, w_2/t had significantly statistical differences between the group of sub-health fatigue and the disease fatigue ($P < 0.05$, $P < 0.01$). The main characteristic of result was that the group of sub-health fatigue and disease fatigue showed a gradual increasing tendency in each parameter compared with the health controls, and it reflected that the two groups of fatigue people had a consistent tendency in the changing nature of pulse. In addition, the changes of pulse feature in the group of disease fatigue were more significant than the sub-health fatigue.

3.4 Results of Modeling and Model Evaluation

Logistic regression method was used to establish classification model based on tongue and pulse data of subjects from the 3 groups (250 in the health controls, 242 in the group of sub-health fatigue and 215 in the disease fatigue) who had complete Body Mass Index (BMI) data. Firstly, multiple logistic regression analysis was used for classification of the three groups based on tongue data, pulse data and BMI data. Table 5 showed the result of classification.

Table 5
Classification table of multiple logistic regression model

Observed	Predicted			
	Healthy controls	sub-health fatigue	disease fatigue	Percent Correct
Healthy controls	167	56	27	66.8%
Sub-health fatigue	77	117	48	48.3%
Disease fatigue	28	30	157	73.0%
Overall Percentage	38.5%	28.7%	32.8%	62.4%

And then to do binary logistic regression analysis. Table 6 showed the classification model result between the group of disease fatigue, the sub-health fatigue and the healthy controls.

Table 6
Classification result of model based on tongue & pulse data

Model	AUC	Accuracy	Sensitivity	Specificity
Healthy controls & sub-health fatigue	0.678	65.70%	73.60%	57.44%
Sub-health & disease fatigue	0.759	67.40%	73.55%	60.47%
Healthy controls & disease fatigue	0.847	76.30%	80.40%	71.63%

The ROC curves were shown in Fig. 7.

In addition, the classification model was reconstructed adding BMI data with tongue & pulse. Table 7showed the classification result of the reconstructed model.

Table 7
Classification result of model based on tongue & pulse & BMI

Model	AUC	Accuracy	Sensitivity	Specificity
Healthy controls & sub-health fatigue	0.678	65.70%	73.60%	57.44%
Sub-health & disease fatigue	0.834	65.10%	78.93%	70.70%
Healthy controls & disease fatigue	0.879	78.90%	82.40%	74.88%

The ROC curves were shown in Fig. 8.

The research result showed that objective data of tongue & pulse had a good classification effect on disease fatigue, followed by sub-health fatigue. After adding BMI data, both of the model accuracy and ROC curve were improved except the sub-health fatigue and healthy controls. BMI is a convenience and noninvasive data, which suggested that we could combine BMI with tongue and pulse data to improve the diagnostic accuracy of fatigue.

Discussion

In this study, the distribution trends of the objective data of tongue were different between the sub-health fatigue population and the disease fatigue population. The study showed that TB-B, TB-R, TB-G, TC-B, TB-I, TB-Y, TB-L, TB-Cb, TB-Cr, TB-a and TB-a were in an ascending order in the group of sub-health fatigue, healthy controls and the group of disease fatigue. This indicated that disease fatigue people in general had more purple or red purple tongue body, and more white-greasy tongue coating. The tongue parameters of sub-health fatigue population were lower than those of the healthy controls, while disease fatigue was higher than the healthy controls, this result might partly related to the fact that the subjects in this study came from the physical examination center. Certain differences were found in tongue parameters of the fatigue groups comparing with the healthy controls, that was subjects in the group of disease fatigue had darker tongue body, more yellow or yellowish brown tongue coating, which was more associated with sthenia syndrome, and the subjects in the group of sub-health fatigue had light-colored tongue with white coating, which was more associated with the deficiency syndrome. The finding was consistent with the TCM theory that sub-health was manifested as decreased vitality, function and adaptability, and disease was mostly due to the hyperactivity of evil spirits, or dysfunction of the dysfunctional organs caused by phlegm, dampness and blood stasis and other pathological products. The result could help to distinguish sub-health fatigue and disease fatigue.

In our study, the pulse analysis result of the three groups showed that fatigue state can directly affect the changes of sphygmogram parameters, and the change had a consistent trend, so to say, the indexes of disease fatigue were more abnormal and the differences were more significant compared to healthy controls, while between the group of sub-health fatigue and health controls only w_2/t had statistical difference, several indexes had significant difference between the group of sub-health fatigue and disease fatigue. As to the distribution trend of pulse indexes, the group of sub-health fatigue was located between healthy controls and the group of disease fatigue. Studies have shown that pulse can directly reflect various cardiovascular functional states, the results of this study, to a certain extent, indicated that patients with disease fatigue had more severe functional decline and other abnormal changes in cardiovascular functions, such as left ventricular function, peripheral resistance, great artery compliance, wall elasticity, blood viscosity. Since fatigue in the most serious case can cause sudden cardiac death, it was of great

practical value to use sphygmogram to detect fatigue in order to diagnose cardiovascular disease and help to guide the early intervention.

BMI is an index of obesity which is closely related to health state. Studies have shown that BMI combined circumference level can be used to assess the risk of coronary heart disease in Japanese diabetic patients[14]. In this study, the accuracy of the model remained unchanged after adding BMI into the group of sub-health fatigue and the healthy controls. The reason may be that there was no difference in BMI index between the two groups, so their contribution to the accuracy of the model was not significant. However, the patients with disease fatigue and sub-health fatigue, there were statistically significant differences in BMI, therefore, BMI combined with tongue and pulse data had a positive effect on the modeling.

Conclusion

In this study, we successfully analyzed the tongue and pulse data characteristics and distribution trend of fatigue and healthy population, at the same time, logistic regression modeling can realize the diagnosis of disease fatigue and sub-health fatigue to a certain extent. It provided a non-invasive differential diagnosis method for the data-driven evaluation of different fatigue states based on the data of tongue and pulse.

Declarations

Acknowledgments

The authors are especially thankful for the positive support received from the Medical Examination Center of Shuguang Hospital affiliated to Shanghai University of Traditional Chinese Medicine and all medical staff involved.

Funding

This research was funded by the National Key Research and Development Program of China (2017YFC1703301), the National Natural Science Foundation of China (81873235, 81973750, 81904094), and 1226 Major Project (BWS17J028). They were not involved in preparing this manuscript or in the decision to submit it for publication. The funder had no role in the study's design, collection, analysis, and interpretation of data, or writing the manuscript.

Compliance with ethical standards

The IRB approved the study protocol of Shuguang Hospital affiliated with Shanghai University of TCM (No. 2018-626-55-01). Written informed consent was obtained from all patients.

Conflict of interest

The authors declare that they have no competing interests.

Availability of data and materials

The datasets generated and analyzed during the current study are not publicly available due to the confidentiality of the data, which is an important component of the National Key Technology R&D Program of the 13th Five-Year Plan (no. 2017YFC1703301) in China, but are available from the corresponding author on reasonable request.

Authors' contributions

Yu-lin Shi and Tao Jiang drafted the initial manuscript, Xiao-juan Hu assisted with data statistical analysis, Li-ping Tu contributed to writing and revising process, and all other authors have assisted in data collection of tongue and pulse, Jia-tuo Xu and Jing-bin Huang contributed to the schemes design and guidance. All authors read and approved the final manuscript.

References

1. Bucur, E, Danet, A.F., Lehr, C.B., Lehr, E., Nita-Lazar, M. (2017) Binary logistic regression-Instrument for assessing museum indoor air impact on exhibits. Journal of the Air & Waste Management Association (1995) 67:391-401. doi:10.1080/10962247.2016.1231724

2. Chen, Y., Chen, Y.B., Tao, R.F. (2017) Interpretation of “Guideline for Prevention and Treatment of Dyslipidemia in Chinese Adults in 2016. Chinese Journal of Practical Internal Medicine 37:38-42

3. Chung, K.F., Yu, Y.M., Yeung, W.F. (2015) Correlates of residual fatigue in patients with major depressive disorder: The role of psychotropic medication. Journal of affective disorders 186:192-197. doi:10.1016/j.jad.2015.07.026

4. Ebede, C.C., Jang, Y., Escalante, C.P. (2017) Cancer-Related Fatigue in Cancer Survivorship. The Medical clinics of North America 101:1085-1097. doi:10.1016/j.mcna.2017.06.007

5. Fatty liver and alcoholic liver Disease Group CSoH, Chinese Medical Association (2001) Diagnostic criteria for nonalcoholic fatty liver Disease (Draft). Chinese journal of hepatology:325

6. Golabi, P., Sayiner, M., Bush, H., Gerber, L.H., Younossi, Z. M. (2017) Patient-Reported Outcomes and Fatigue in Patients with Chronic Hepatitis C Infection. Clinics in liver disease 21:565-578. doi:10.1016/j.cld.2017.03.011

7. Guidelines for the prevention and control of type 2 diabetes in China. (2018). Chinese journal of practical internal medicine 38:292-344

8. Zhu, H.H., Zhang, Z.F.,Chen, X., Fei, Z.F., Xu, J.T. (2010) .Design and evaluation of the Simple Health Assessment Questionnaire H2O. In: The fourth National Symposium on Integrated Chinese and Western Medicine diagnosis, Hohhot, Inner Mongolia, China, 2010. p 4

9. Hu, M.C., Lan, K.C., Fang, W. C, Huang, Y.C., Ho, T. J., Lin, C. P., Yeh, M. H., Raknim, P., Lin, Y. H., Cheng, M. H., He, Y.T., Tseng, K. C. (2019) .Automated tongue diagnosis on the smartphone and its applications. Computer methods and programs in biomedicine 174:51-64. doi:10.1016/j.cmpb.2017.12.029

10. Hu, X.J., Zhang, L., Xu, J.T., Liu, B.C., Wang, J.Y., Hong, Y.L., Tu, L.P., Cui, J. (2018) .Pulse Wave Cycle Features Analysis of Different Blood Pressure Grades in the Elderly. Evidence-based complementary and alternative medicine : eCAM 2018:1976041. doi:10.1155/2018/1976041

11. Cui, J., Tu, L.P., Zhang, J.F., Hu,X.J., Huang, J.B., Wang, G.L., Wang, W.J., Zhang, Z.F., Xu,J.T. (2018). Research on pulse graph characteristics of 1 720 cases with different health status and age gradient. Shanghai Journal of Traditional Chinese Medicine 52:15-23

12. Zhang, J.F., Xu, J.T., Tu, L.P., Zhang, Z.F., Hu, X.J., Cui,J., Huang, J.B., Cui, L.T., Zhang, S.L., Jing, C.C. (2017). Study on the Characteristics of Sub-health Symptoms and TCM Syndrome Patterns Distribution in 1 754 Non-disease Population Chinese Journal of Integrative Medicine 37:934-938

13. Huang, J.B., Xu, J.T., Zhang,Z.F., Tu, L.P., Cui,J., Hu, X.J., Cui,L.T., Jiang, T. (2018) .Influence of 'daily rhythm' factors on tongue image characteristics of healthy people. China Journal of Traditional Chinese Medicine and Pharmacy 33:3462-3465

14. Komatsu, T., Fujihara, K., Yamada, M.H., Sato, T., Kitazawa, M., Yamamoto, M., Seida, H., Yamanaka, N., Matsubayashi, Y., Yamada, T., Sone, H. (2020). 449-P: Impact of Body Mass Index (BMI) and Waist Circumference (WC) on Coronary Artery Disease (CAD) in Japanese with and without Diabetes Mellitus (DM). Diabetes 69

15. Kung, Y.Y., Kuo, T.B.J., Lai, C.T., Shen, Y.C., Su,Y.C., Yang, C.C.H. (2021). Disclosure of suboptimal health status through traditional Chinese medicine-based body constitution and pulse patterns. Complementary therapies in medicine 56:102607. doi:10.1016/j.ctim.2020.102607

16. Li, J., Yuan, P., Hu, X.J, Huang, J.B., Cui, L.T., Cui, J., Ma, X.X., Jiang, T., Yao, X.H., Li, J.C., Shi, Y.L., Bi, Z.J., Wang, Y., Fu, H.Y., Wang, J., Lin, Y.T., Pai, C.H., Guo, X.J., Zhou, C.L., Tu, L.P., Xu, J.T. (2021). A tongue features fusion approach to predicting prediabetes and diabetes with machine learning. Journal of biomedical informatics 115:103693. doi:10.1016/j.jbi.2021.103693

17. Wang, L.L., Zhang,X,Y., Peng,M. (2019). Objective analysis of complexion and tongue color in patients with chronic fatigue syndrome. . Shandong Medical Journal 59:81-83. doi:10.1016/j.jad.2015.07.026

18. Liu, J. (2020) .Highlights of the 2018 Chinese hypertension guidelines. Clinical hypertension 26:8. doi:10.1186/s40885-020-00141-3

19. Luo, Z.Y., Cui, J., Hu, X.J., Tu, L.P., Liu, H.D., Jiao, W., Zeng, L.Z., Jing, C.C., Qiao, L.J., Ma, X.X., Wang, Y., Wang, J., Pai, C.H., Qi, Z., Zhang, Z.F., Xu, J.T. (2018). A Study of Machine-Learning Classifiers for Hypertension Based on Radial Pulse Wave. BioMed research international 2018:2964816. doi:10.1155/2018/2964816

20. Persson, P.B., Bondke, Persson. A. (2016).Fatigue. Acta physiologica (Oxford, England) 218:3-4. doi:10.1111/apha.12756

21. Shi, H.Z., Fan, Q.C., Gao, J.Y., Liu, J.L., Bai, G.E., Mi, T., Zhao, S., Liu, Y., Xu, D., Guo, Z.F., Li, Y.Z. (2017). Evaluation of the health status of six volunteers from the Mars 500 project using pulse analysis. Chinese journal of integrative medicine 23:574-580. doi:10.1007/s11655-016-2539-5

22. Swain, M.G., Jones, D.E.J. (2019). Fatigue in chronic liver disease: New insights and therapeutic approaches. Liver international : official journal of the International Association for the Study of the Liver 39:6-19. doi:10.1111/liv.13919

23. Tang, A.C., Chung, J.W., Wong, T.K. (2012) .Digitalizing traditional chinese medicine pulse diagnosis with artificial neural network. Telemedicine journal and e-health : the official journal of the American Telemedicine Association 18:446-453. doi:10.1089/tmj.2011.0204

24. Thirunavukkarasu, U., Umapathy, S., Krishnan, P.T., Janardanan, K. (2020). Human Tongue Thermography Could Be a Prognostic Tool for Prescreening the Type II Diabetes Mellitus. Evidence-based complementary and alternative

medicine : eCAM 2020:3186208. doi:10.1155/2020/3186208

25. Zhang, B., Wang, X., You, J., Zhang, D. (2013) .Tongue color analysis for medical application. Evidence-based complementary and alternative medicine : eCAM 2013:264742. doi:10.1155/2013/264742

26. Zhang, J.F., Xu, J.T., Hu, X.J., Chen, Q.G., Tu, L.P., Huang, J.B., Cui, J. (2017). Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images. BioMed research international 2017:7961494. doi:10.1155/2017/7961494

27. Zhang, K., Geng, W., Zhang, S. (2018). Network-based logistic regression integration method for biomarker identification. BMC systems biology 12:135. doi:10.1186/s12918-018-0657-8

Figures

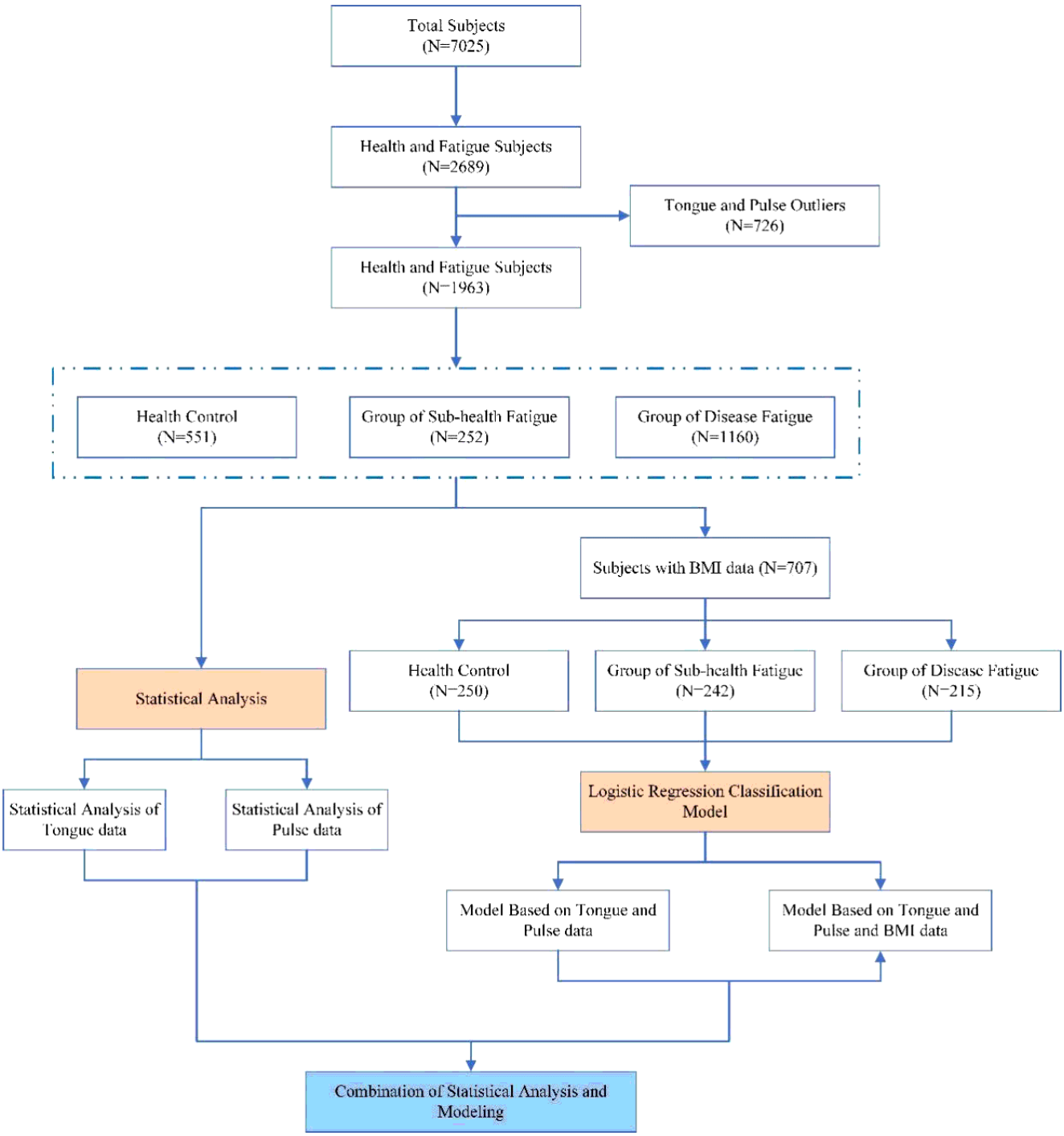

Figure 1

Overall Flow Diagram

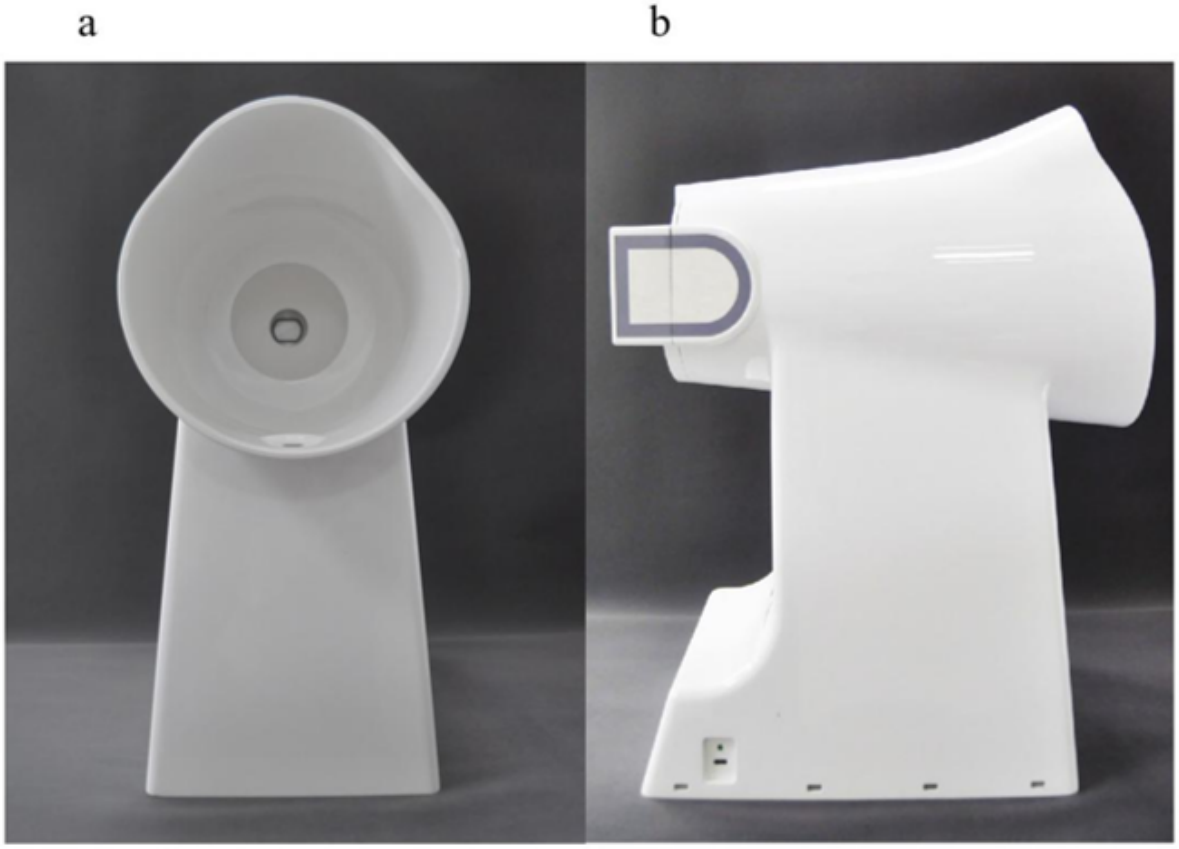

Figure 2

TFDA-1 tongue and face diagnosis instrument. a: Front view. b: Profile view.

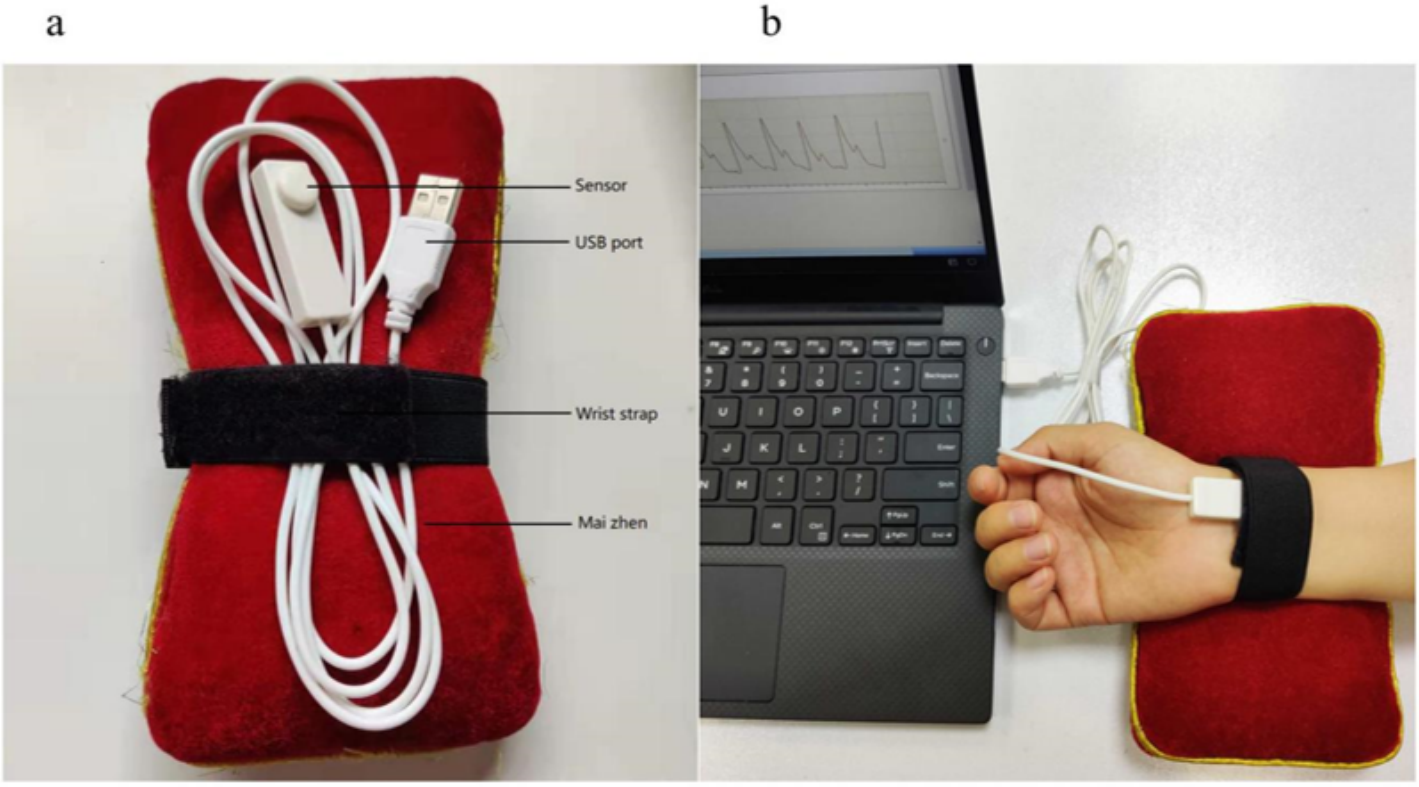

Figure 3

PDA-1 pulse diagnosis instrument and corresponding collection picture. a: PDA-1 pulse diagnosis instrument; b: Real picture of pulse acquisition

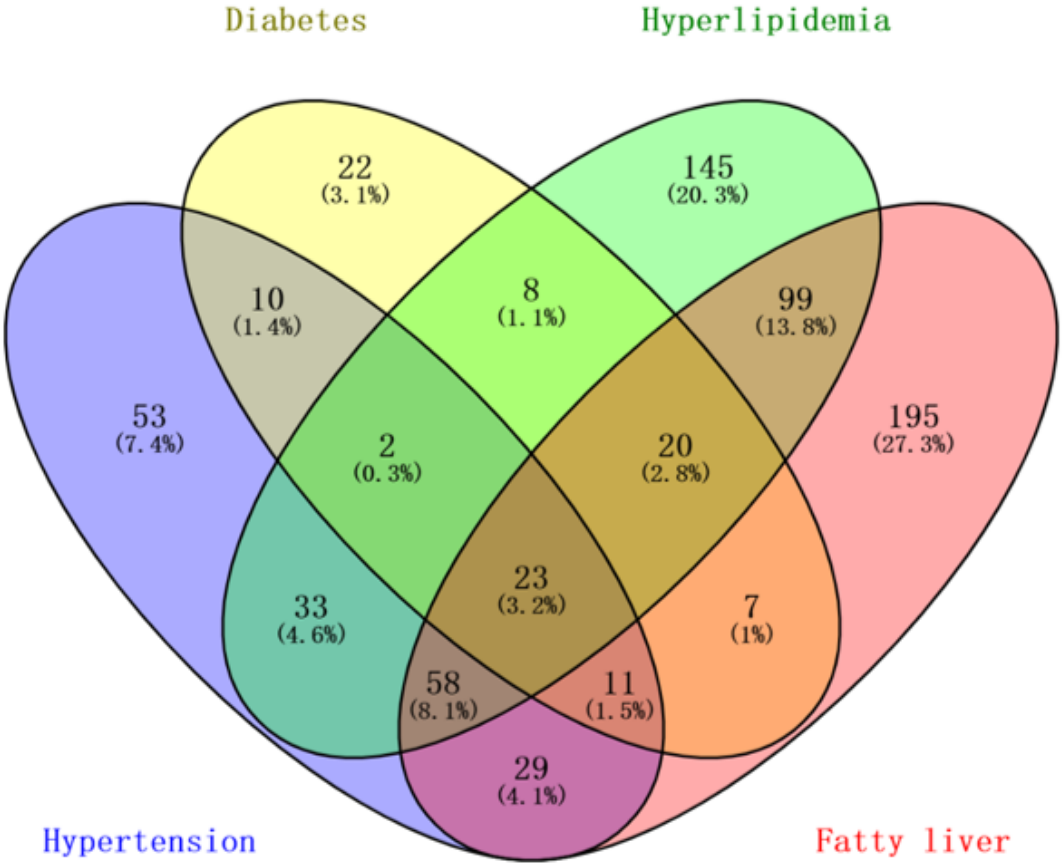

Figure 4

Distribution of main diseases in the group of disease fatigue

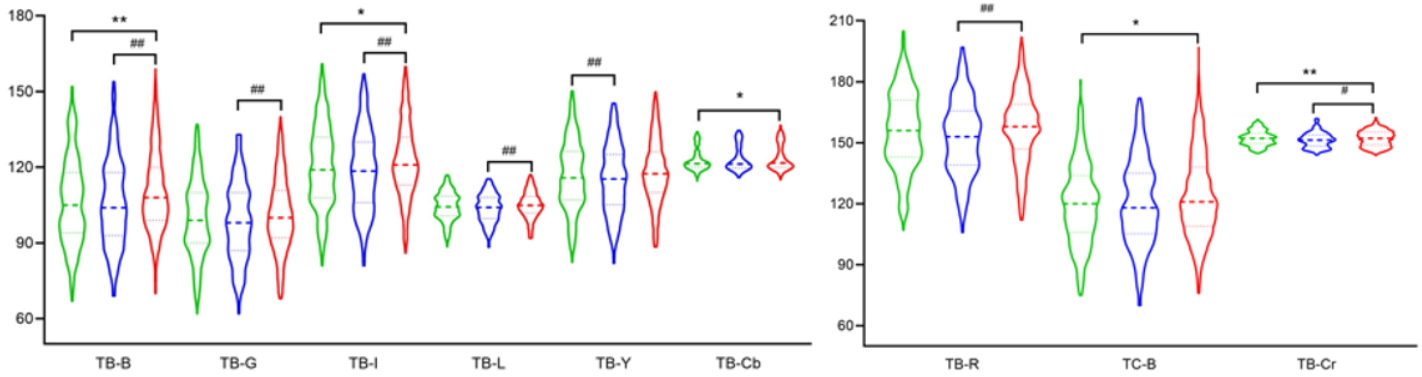

Figure 5

Violin Plots of the tongue characteristic parameters of the three groups

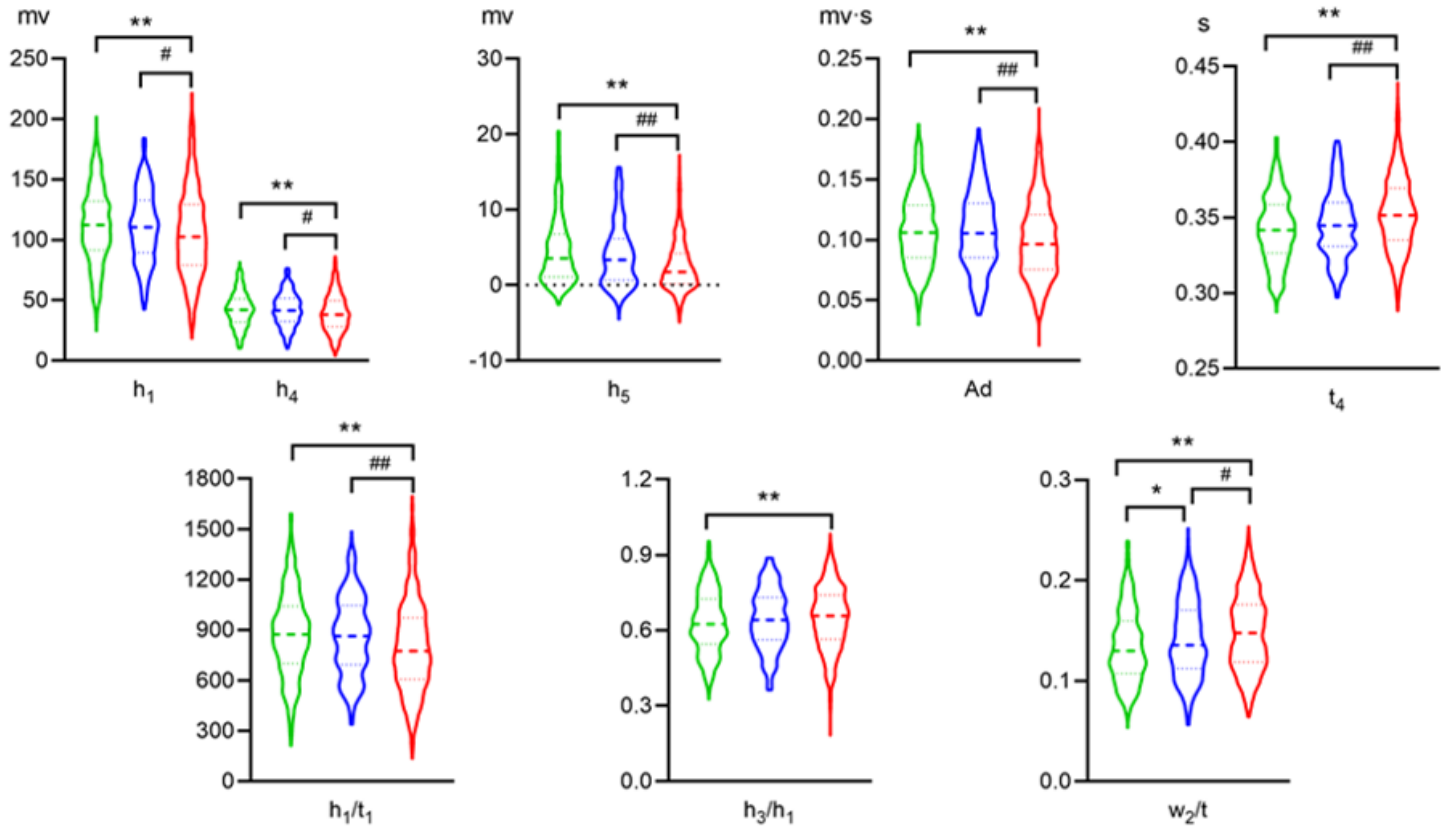

Figure 6

Violin Plots of the pulse characteristic parameters of the three groups

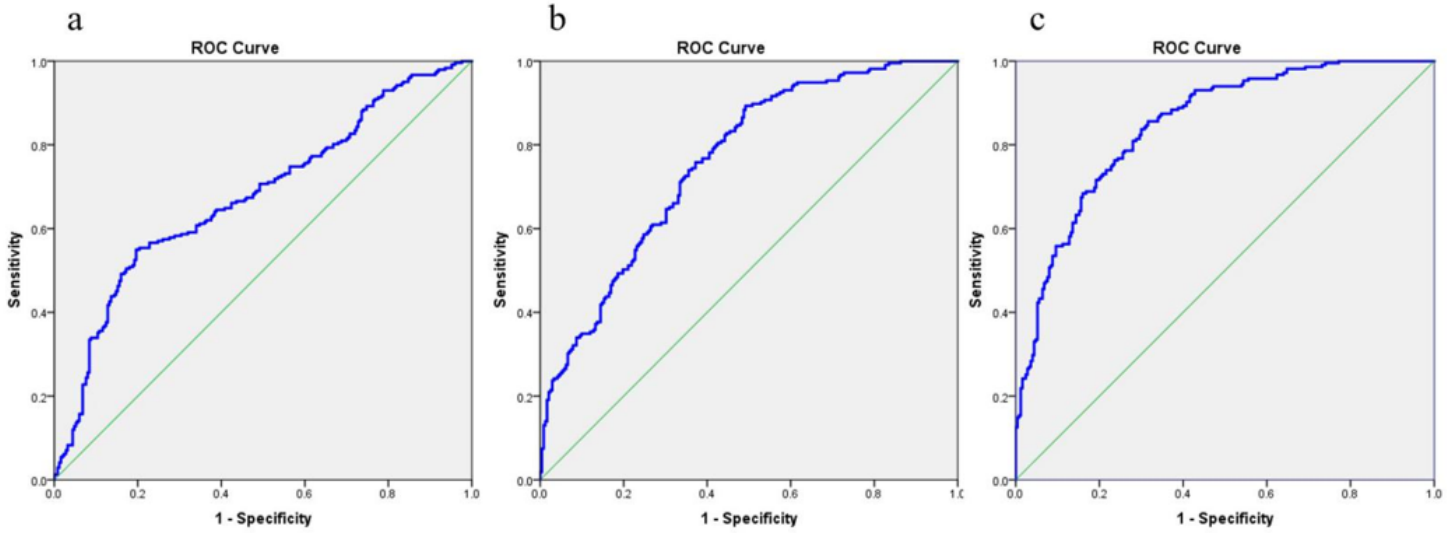

Figure 7

ROC curves of classification model based on tongue & pulse. a: Health & sub-health fatigue; b: Sub-health & disease fatigue; c: Health & disease fatigue

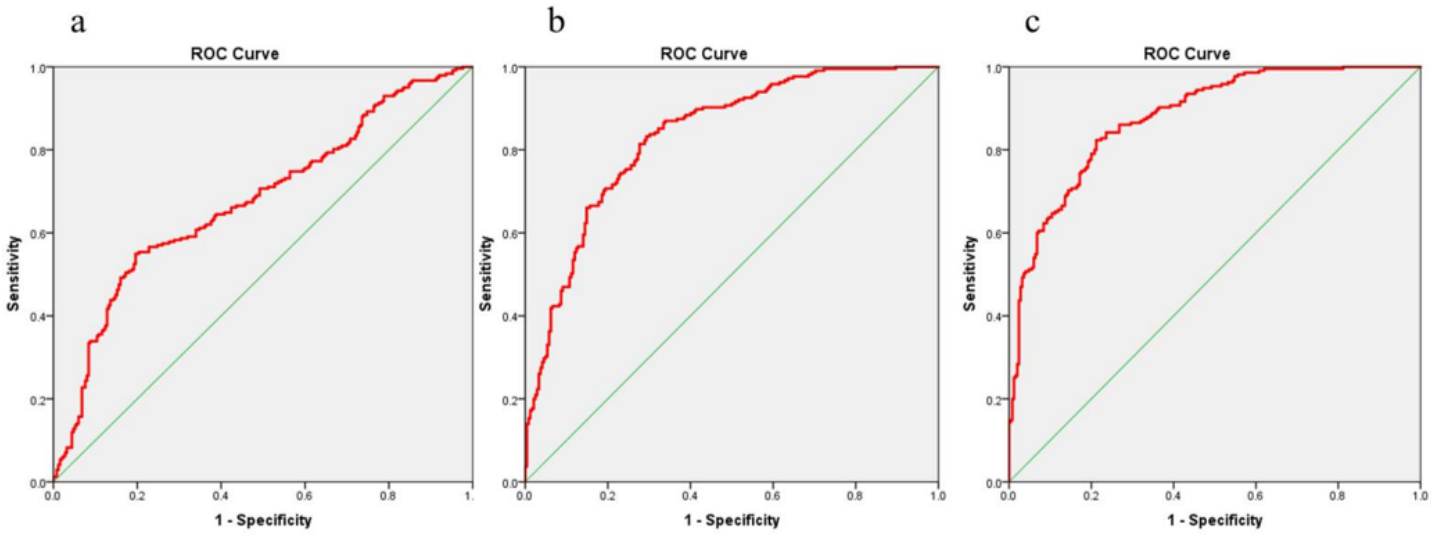

Figure 8

ROC curves of classification model based on tongue & pulse & BMI. a: Health & sub-health fatigue; b: Sub-health & disease fatigue; c: Health & disease fatigue