Study on the Syndrome Characteristics and Classification Model of Non-Small Cell Lung Cancer Based on Tongue and Pulse Data

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

Jiayi Liu
Shanghai University of Traditional Chinese Medicine

Xiaojuan Hu
Shanghai University of Traditional Chinese Medicine

Liping Tu
Shanghai University of Traditional Chinese Medicine

Ji Cui
Shanghai University of Traditional Chinese Medicine

Jun Li
Shanghai University of Traditional Chinese Medicine

Zijuan Bi
Shanghai University of Traditional Chinese Medicine

Jiacai Li
Shanghai University of Traditional Chinese Medicine

Ling Xu
Shanghai University of Traditional Chinese Medicine Yueyang Hospital of Integrated Traditional Chinese Medicine and Western Medicine

Jiatuo Xu
Shanghai University of Traditional Chinese Medicine

Research

Keywords: non-small cell lung cancer, TCM-syndrome, tongue and pulse data, classification model, machine learning

DOI: https://doi.org/10.21203/rs.3.rs-355613/v1

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

**Background:** Lung cancer is a common malignant tumor that affects people's health seriously. Traditional Chinese medicine (TCM) is one of the effective methods for the treatment of advanced lung cancer; accurate TCM syndrome differentiation is essential to treatment. When the symptoms are not obvious, the traditional symptom-based syndrome differentiation cannot be carried out. There is a close relationship between syndrome and index of western medicine, the combination of micro index and macro symptom can assist syndrome differentiation effectively.

**Methods:** Tongue and pulse data of non-small cell lung cancer (NSCLC) patients with Qi deficiency syndrome (n=163), patients with Yin deficiency syndrome (n=174) and healthy controls (n=185) were collected by using intelligent Tongue and Face Diagnosis Analysis-1 instrument and Pulse Diagnosis Analysis-1 instrument, respectively. The characteristics of tongue and pulse data were analyzed, the correlation analysis was also made on tongue and pulse data. And four machine learning methods, namely Random Forest, Logistic Regression, Support Vector Machine and Neural Network, were used to establish the classification models based on symptoms, tongue & pulse data, and symptoms & tongue & pulse data, respectively.

**Results:** Significant difference indexes of tongue diagnosis between Qi deficiency syndrome and Yin deficiency syndrome were TB-a, TB-S, TB-Cr, TC-a, TC-S, TC-Cr, perAll and the tongue coating texture indexes including TC-Con, TC-ASM, TC-MEAN, and TC-ENT. Significant difference indexes of pulse diagnosis were t₄ and t₅. The classification performance of each model based on different data sets was as follows: model of tongue & pulse data < model of symptom < model of symptom & tongue & pulse data. The Neural Network model had a better classification performance for the symptom & tongue & pulse data, with an area under the ROC curve and accuracy rate were 0.9401 and 0.8806.

**Conclusions:** This study explored the characteristics of tongue and pulse data of NSCLC with Qi deficiency syndrome and Yin deficiency syndrome, and established syndromes classification model. It was feasible to use tongue and pulse data as one of the objective diagnostic indexes in Qi deficiency syndrome and Yin deficiency syndrome of NSCLC.

Name of the registry: Chinese Clinical Trial Registry

Trial registration number: ChiCTR1900026008; ChiCTR-IOR-15006502

Date of registration: Jun. 04th, 2015

URL of trial registry record: http://www.chictr.org.cn/showprojen.aspx?proj=11119;
http://www.chictr.org.cn/edit.aspx?pid=38828&htm=4 (This is a retrospective registration)

Introduction

Lung cancer is a common malignant tumor of the lung, which is one of the highest morbidity and mortality in the world. It is estimated that the number of deaths from lung cancer accounts for about 24% of all cancer deaths in the United States[^13]. Organization report shows that lung cancer causes approximately 1.76 million deaths worldwide each year, accounting for 18.7% of all cancer deaths[^3]. Non-small cell lung cancer (NSCLC) is the most common histological type of lung cancer, accounting for more than 80% of primary lung cancers, with a very high morbidity and mortality[^4], 60% of NSCLC cases have metastasized at the time of diagnosis, and the 5-year survival rate for advanced NSCLC is less than 5%, and early diagnosis of lung cancer is an important opportunity to reduce mortality[^5,6]. The current treatment methods for NSCLC mainly include surgery, radiotherapy, chemotherapy and targeted therapy[^7,8], chemotherapy is the most common treatment, patients with poor health often have a low tolerance to conventional treatment, and they often tend to drug resistance[^9]. Traditional Chinese medicine (TCM) has a long history and rich experience in the treatment of lung cancer, and is one of the main methods of comprehensive treatment of lung cancer in China. Systematic evaluation of TCM showed that TCM combined with radiotherapy and chemotherapy and targeted therapy had certain advantages in alleviating symptoms, stabilizing tumors, improving life quality and prolonging survival period[^10]. TCM has been proved to be an effective method for the treatment of advanced lung cancer. On the basis of accurate syndrome differentiation, TCM plays an active role in each stage of the occurrence and development of lung cancer[^11,12].

Syndrome differentiation and treatment is the basic principle of TCM to diagnose and deal with diseases. It is a process of comprehensive judgment on the four diagnostic information of patients based on the theory of TCM combined with the doctor's experience[^13]. Accurate syndrome differentiation can provide a basis for the treatment of diseases and is the foundation of clinical efficacy. Traditional syndrome differentiation and treatment inevitably suffer from subjectivity and ambiguity, which actually hinders the development of TCM. Micro-syndrome differentiation is a method of using modern advanced technology to go deep into the body's microcosmic level to understand and differentiate syndromes on the basis of macroscopic syndrome differentiation, which can be used to guide disease differentiation and
syndrome differentiation, explore the cause and pathogenesis, and evaluate the efficacy and guide the prognosis of the disease\[^{14}\]. Previous studies have verified that there is a close relationship between different syndromes and physical and chemical indexes, and the combination of micro index and macro symptom can assist syndrome differentiation effectively.

With the rapid development of modern research on tongue and pulse diagnosis, a variety of tongue and pulse diagnosis instruments are widely used in clinical practice, generating a large number of objective data of tongue and pulse diagnosis, which are also microscopic indexes in a sense. In recent years, researches based on tongue and pulse diagnosis data have been increasing day by day, many researchers apply machine learning and data mining methods to the fields of image recognition, target detection, natural language processing and other fields\[^{15–18}\]. In addition, researches have demonstrated that accurate detection, identification and multi-dimensional quantitative analysis based on tongue and pulse data have been gradually applied to disease diagnosis. By constructing the diagnostic relationship between tongue and pulse and health status, it not only saves medical resources, but also greatly improves diagnosis efficiency and treatment\[^{19–22}\].

Qi deficiency syndrome and Yin deficiency syndrome are the two main common syndromes of NSCLC. When the symptoms are not obvious, the traditional symptom-based syndrome differentiation cannot be carried out. Modern study of tongue and pulse diagnosis research provides a good data basis for TCM syndrome differentiation. Therefore, this study aims to explore the differentiation of NSCLC based on tongue & pulse diagnosis data, using machine learning methods to establish a syndrome classification model based on macro symptom data, objective tongue & pulse data, and macro symptom data & objective tongue & pulse data, and evaluate the contribution rate of the objective data of tongue and pulse diagnosis to syndrome differentiation.

Materials And Methods

Study design and subjects

A total of 337 patients were selected from the Oncology Department of Yueyang Hospital of Integrated Traditional Chinese and Western Medicine from January 2018 to October 2020, including 163 patients with Qi deficiency syndrome and 174 patients with Yin deficiency syndrome, all patients were pathologically or cytologically confirmed to be NSCLC. A total of 184 healthy populations were randomly selected from Shuguang Hospital of Shanghai University of Traditional Chinese Medicine from January 2018 to October 2020 as the healthy controls. The research flowchart was shown in Fig. 1.

Diagnostic criteria

Diagnostic criteria of Western medicine: according to the clinical practice guidelines for lung cancer screening issued by the National Comprehensive Cancer Network (NCCN)\[^{23}\] and the fourth edition lung cancer histological classification standards of “Classification of Lung Tumors”\[^{24,25}\] issued by the World Health Organization (WHO).

TCM Syndrome Differentiation Standard: according to the “Technical Guidelines for Clinical Research of New Drugs of Syndromes”\[^{26}\] and the Syndrome Part of TCM Clinical Diagnosis and Treatment Terms\[^{27}\] and textbooks of Common Diseases and Symptoms in Internal Medicine of Traditional Chinese Medicine.

Inclusion And Exclusion Criteria

Inclusion criteria: (1) Meet the above diagnostic criteria. (2) Confirmed by pathology or cytology. (3) No serious liver or kidney damage. (4) Know and sign informed consent.

Exclusion criteria: (1) Those who do not meet the inclusion criteria for NSCLC. (2) Patients with severe primary diseases such as cardiovascular, cerebrovascular, liver, kidney and blood system. (3) Pregnant or lactating women. (4) Psychopath. (5) Patients who are unable to cooperate with research work due to subjective and objective reasons and who have poor compliance.

Collecting Clinical Tongue And Pulse Data

TFDA-1 digital tongue and face diagnosis instrument and PDA-1 digital pulse diagnosis instrument developed by the National Key Research and Development Program were respectively used to collect the tongue and pulse diagnostic data of the patients. The Information Record Form of TCM Clinical Four Diagnostics (Copyright No.: 2016Z11L025702) of our research group was used to record the symptoms of patients\[^{28}\]. All the work of tongue and pulse diagnosis collection and inquiry is completed by professional personnel of TCM or integrated TCM and western medicine who have received standardized training. Each patient was consulted by at least two professional researchers,
and the interpretation of all patient syndromes was completed by three senior doctors to ensure the consistency and authenticity of data collection and interpretation, and minimize deviation. The TFDA-1 digital tongue diagnosis instrument and its corresponding tongue image analysis system were shown in Fig. 2.

The PDA-1 digital pulse diagnosis instrument and its corresponding sphygmogram were shown in Fig. 3.

Tongue diagnosis parameters mainly come from the three color spaces of Lab, HIS and YCrCb\cite{29–32}, and each parameter of tongue and pulse diagnosis has its corresponding medical significance\cite{31, 33}.

### Statistical analysis

SPSS 25.0 was used for statistical analysis. Count data was expressed as percentage N(%), measurement data obeyed normal distribution was expressed as “mean ± standard deviation”, and those didn’t obey is expressed as “median (upper quartile, lower quartile)”. Measured data were compared with analysis of variance (ANOVA) or rank sum test (Kruskal-Wallis H test), the correlation heat map was made by GraphPad Prism 8.0. All the test results were double-tailed test, and the test level $\alpha = 0.05$, the difference was statistically significant when $P < 0.05$.

### Classification By Machine Learning Approach

Orange (3.26.0) software was used, four machine learning methods, namely Neural Network, Random Forest, Support Vector Machine (SVM) and Logistic Regression were used to set the ratio of training set to test set at 8:2. Adjusted the parameters of each model to establish classification and diagnosis models of Qi deficiency syndrome and Yin-deficiency syndrome of NSCLC based on “symptom”, “tongue pulse” and “symptom & tongue pulse” respectively. Accuracy, Precision, F1-score (F1), Sensitivity, Specificity and AUC were used as evaluation indexes to evaluate the predictive performance. AUC was the area under the ROC Curve, and the value was between 0.5-1. The larger the value, the better the classification effect of the classifier. The calculation formula of each index was as follows:

\[
\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \\
\text{Precision} = \frac{TP}{TP + FP} \times 100\% \\
\text{Sensitivity} = \frac{TP}{TP + FN} \times 100\% \\
\text{Specificity} = \frac{TN}{TN + FP} \times 100\% \\
\text{F1} = \frac{2 \times \text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}}
\]

In the above statements, True Positive (TP) was the positive sample predicted by the model as the positive category. True Negative (TN) was the negative sample predicted by the model as the negative category. False Positives (FP) was the negative sample predicted by the model as the positive category. False Negative (FN) was the positive sample predicted by the model as the negative category.

### Results

#### Characteristics of participants

The basic statistical analysis result of the three groups was shown in Table 1.

| Characteristic         | Healthy Controls (n = 184) | Qi deficiency syndrome (n = 163) | Yin deficiency syndrome (n = 174) |
|------------------------|----------------------------|---------------------------------|-----------------------------------|
| Sex, n(%)              | Male 96(52.17)             | 72(44.17)                       | 89(51.15)                         |
|                        | Female 88(47.83)           | 91(55.83)                       | 85(48.85)                         |
| Age, years             | 27.00(29.00-24.25)         | 67.00(59.00–71.00)**            | 67.00(60.00–72.00)**              |
| ** vs. Healthy Controls, P < 0.01** | | | |

Page 4/19
The result showed that: in terms of sex ratio, the female in the syndrome of Qi deficiency was slightly higher than the male, and the male in the syndrome of Yin deficiency was slightly higher than the female. Compared with the healthy controls, the age of the people with Qi deficiency syndrome and Yin deficiency syndrome were statistically significant, while the age difference between Qi deficiency syndrome and Yin deficiency syndrome was not statistically significant.

### Statistical Analysis Of Tongue Data

The statistical analysis result of tongue diagnosis data in the three groups was shown in Table 2.

| Domain | Color space | Index | Healthy Controls (n = 184) | Qi deficiency syndrome (n = 163) | Yin deficiency syndrome (n = 174) |
|--------|-------------|-------|-----------------------------|----------------------------------|----------------------------------|
| TB     | Lab         | TB-L  | 103.99 (100.81-108.79)      | 96.31 (75.15-102.89)**           | 99.83 (80.51-103.24)**           |
|        |             | TB-a  | 19.98 ± 2.82                | 19.31 ± 3.81                     | 21.06 ± 4.23**###               |
|        |             | TB-b  | 4.76 (0.82-7.00)            | 7.04 (5.47–8.28)**              | 7.04 (5.47–8.28)**              |
| HIS    |             | TB-H  | 176.22 (168.50-180.95)      | 180.00 (177.98-182.83)**         | 180.00 (177.98-182.83)**         |
|        |             | TB-S  | 0.17 (0.16–0.20)            | 0.17 (0.15–0.19)                 | 0.17 (0.15–0.19)**##            |
|        |             | TB-I  | 117.00 (108.00-132.00)      | 116.00 (109.00-126.00)**         | 116.00 (109.00-126.00)          |
| YCrCb  |             | TB-Y  | 114.98 (107.03-126.56)      | 114.35 (106.90-123.72)           | 114.35 (106.90-123.72)*         |
|        |             | TB-Cr | 151.41 ± 3.05               | 152.29 ± 3.89                    | 154.15 ± 4.44**##               |
|        |             | TB-Cb | 121.61 (119.75-124.82)      | 119.84 (118.53-120.99)**         | 119.27 (118.09-120.57)**        |
| texture index |       | TB-Con | 71.47 (46.96–99.54) | 74.56 (48.28–94.64) | 60.96 (45.32–86.08) |
| TC     | Lab         | TC-L  | 109.24 (104.97-113.54)      | 89.38 (76.22-104.87)**           | 95.35 (82.53-105.08)**           |
|        |             | TC-a  | 12.31 ± 2.69                | 12.75 ± 3.21                     | 14.25 ± 3.78**###               |
|        |             | TC-b  | 2.71 (-1.16-5.32)           | 5.59 (4.24–6.62)**               | 5.86 (4.35–7.26)**              |
| HIS    |             | TC-H  | 176.70 (162.43-183.25)      | 183.00 (180.00-186.35)**         | 182.58 (178.64-185.72)**         |
|        |             | TC-S  | 0.11 (0.09–0.13)            | 0.12 (0.10–0.14)*                | 0.13 (0.11–0.17)**##            |
|        |             | TC-I  | 130.00 (117.00-142.75)      | 119.00 (99.00-135.00)**          | 115.00 (92.75–133.00)**          |
| YCrCb  |             | TC-Y  | 126.78 (115.63–137.70)      | 118.65 (99.72-132.07)**          | 114.02 (95.19-129.53)**          |
|        |             | TC-Cr | 142.89 (140.89-145.181)     | 143.97 (142.27–146.51)**         | 145.49 (143.00-148.79)**##      |
|        |             | TC-Cb | 123.90 (121.54-127.61)      | 123.36 (120.34-122.81)**         | 121.35 (120.01-122.67)**        |
| TC index | perAll       | 0.54 (0.43–0.69)           | 0.44 (0.34–0.50)**              | 0.38 (0.21–0.50)**##            |
|         | perPart      | 1.09 (1.02–1.22)           | 1.24 (1.11–1.42)**              | 1.28 (1.11–1.57)**##            |
| texture index |       | TC-Con | 89.27 (62.31-124.17) | 83.13 (63.82–123.30) | 71.53 (44.56-115.98)**##         |
|         |            | TC-ASM | 0.07 (0.06–0.08)           | 0.07 (0.06–0.08)                 | 0.08 (0.06–0.10)**##            |
|         |            | TC-MEAN | 0.03 (0.02–0.03)           | 0.03 (0.02–0.03)                 | 0.03 (0.02–0.03)**##            |
|         |            | TC-ENT | 1.26 (1.18–1.34)           | 1.25 (1.18–1.34)                 | 1.21 (1.09–1.31)**##            |

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

#vs. Qi deficiency syndrome P < 0.05, ## vs. Qi deficiency syndrome P < 0.01
The result showed that: (1) Compared with Qi deficiency syndrome, there were more significant differences between Yin deficiency syndrome and the healthy controls. (2) In the indexes of significant difference between Yin deficiency syndrome and healthy controls, except for the texture index of tongue coating, the index of the tongue body is more than that of the tongue coating. That is, changes of tongue body index of Yin deficiency syndrome were more significant than that of tongue coating index. (3) Significant difference indexes of tongue diagnosis between Qi deficiency syndrome and Yin deficiency syndrome were TB-a, TB-S, TB-Cr, TC-a, TC-S, TC-Cr, perAll and the TC-Con, TC-ASM, TC-Mean, TC-ENT of the tongue coating texture indexes. Among them, the parameters of TB-a, TB-Cr, TC-a, TC-S, TC-Cr, and TC-ASM of Yin deficiency syndrome were all higher than those of Qi deficiency syndrome, while the parameters of perAll, TC-Con, TC-ENT of Yin deficiency syndrome were all lower than those of Qi deficiency syndrome.

**Statistical Analysis Of Pulse Data**

The statistical analysis result of pulse diagnosis data in the three groups was shown in Table 3.

| Index | Healthy Controls (n = 184) | Qi deficiency syndrome (n = 163) | Yin deficiency syndrome (n = 174) |
|-------|----------------------------|---------------------------------|----------------------------------|
| t₁(s) | 0.13 (0.12–0.14)           | 0.14 (0.13–0.15)**              | 0.14 (0.13–0.14)**               |
| t₄(s) | 0.34 (0.32–0.36)           | 0.37 (0.35–0.39)**              | 0.37 (0.34–0.39)**#              |
| t₅(s) | 0.41 (0.39–0.42)           | 0.43 (0.41–0.46)**              | 0.42 (0.40–0.44)**##             |
| t(s)  | 0.80 (0.75–0.88)           | 0.86 (0.76–0.97)**              | 0.84 (0.72–0.94)                 |
| h₁(mv) | 13.89 (11.53–16.41)      | 10.99 (7.62–15.42)**           | 11.56 (8.86–16.51)**            |
| h₃(mv) | 8.48 (6.56–10.59)         | 6.64 (4.38–10.07)**            | 7.18 (4.85–10.12)**             |
| h₄(mv) | 5.21 (4.18–6.32)          | 2.18 (1.37–3.24)**             | 2.53 (1.44–3.50)**              |
| h₅(mv) | 0.50 (0.15–0.95)          | 0.23 (0.05–0.69)**             | 0.21 (0.05–0.60)**              |
| h₁/h₁ | 0.62 (0.52–0.70)          | 0.61 (0.53–0.71)               | 0.60 (0.49–0.73)                |
| h₃/t₁ | 4.43 (3.49–5.35)          | 3.22 (2.26–4.57)**             | 3.45 (2.68–4.82)**              |
| h₃/h₁ | 0.38 (0.32–0.43)          | 0.21 (0.12–0.31)**             | 0.21 (0.14–0.28)**              |
| t₁/t  | 0.16 (0.14–0.17)          | 0.16 (0.14–0.19)               | 0.17 (0.14–0.19)                |
| t₄/t₅ | 0.83 (0.80–0.88)          | 0.86 (0.82–0.91)**             | 0.87 (0.82–0.91)**              |
| w₁/t  | 0.20 (0.15–0.23)          | 0.21 (0.19–0.23)**             | 0.21 (0.19–0.23)**              |
| w₂/t  | 0.12 (0.10–0.16)          | 0.15 (0.13–0.18)**             | 0.15 (0.13–0.18)**              |

* vs. Healthy Controls P < 0.05, ** vs. Healthy Controls, P < 0.01
# vs. Qi deficiency syndrome P < 0.05, ## vs. Qi deficiency syndrome P < 0.01

The result showed that: (1) The pulse parameters t₁, t₄, t₅, h₁, h₃, h₄, h₅, h₁/t₁, h₃/h₁, t₄/t₅, w₁/t, w₂/t of Qi deficiency syndrome and Yin deficiency syndrome had statistical significance compared with those of healthy controls. (2) Only two parameters, t₄ and t₅, showed statistically significant differences between Qi deficiency syndrome and Yin deficiency syndrome.

**Correlation Analysis Of Tongue And Pulse Data**

Further correlation analysis was conducted on the tongue and pulse data with statistical significance about Qi deficiency syndrome and Yin deficiency syndrome. The correlation heat map was made by GraphPad Prism 8.0, and all the results were two-tailed tests. The difference was statistically significant when P < 0.05, the heat map result of Qi deficiency syndrome was shown in Fig. 4.
The correlation analysis result of tongue and pulse data between Qi deficiency syndrome was shown in Table 4.

Table 4
Correlation analysis of tongue and pulse diagnosis data of Qi deficiency syndrome

| Index       | perAll | TC-Con | TC-ASM | TC-ENT | TC-MEAN | TB-S | TC-S | TB-a | TC-a | TB-Cr | TC-Cr | t4 | t5 |
|-------------|-------|--------|--------|--------|---------|------|------|------|------|------|-------|----|----|
| perAll      | 1.00  | 0.16*  | -0.14  | 0.16*  | 0.14    | -0.31**| -0.36**| -0.40**| -0.52**| -0.53**| 0.11  | 0.11|
| TC-Con      | 1.00  | 0.16*  | -0.99**| 1.00   | 1.00**  | 0.33**| 0.32**| 0.22**| 0.14  | 0.14  | 0.17* | 0.12|
| TC-ASM      | -0.14 | -0.99**| 1.00   | 0.16** | 1.00    | -0.33**| -0.32**| -0.23**| -0.15 | -0.03  | -0.18*| 0.12|
| TC-ENT      | 0.16* | 1.00** | -0.99**| 1.00   | 1.00    | -0.33**| -0.32**| -0.23**| 0.14  | -0.03  | -0.13 | 0.12|
| TC-MEAN     | 0.14  | 1.00** | 1.00   | 1.00   | 1.00**  | -0.33**| -0.32**| -0.23**| 0.14  | 0.14  | 0.18* | 0.12|
| TB-S        | -0.31**| 0.33**| 0.33** | -0.33**| 0.60**  | 0.32**| 0.21**| 0.51**| 0.80**| 0.84**| 0.58**| 1.00|
| TC-S        | -0.36**| 0.32**| -0.32**| -0.32**| 1.00    | 0.32**| 0.21**| 0.51**| 0.84**| 0.84**| 1.00  | 1.00|
| TB-a        | -0.36**| 0.21**| 0.21** | 0.51** | 0.59**  | 0.22**| 0.21**| 0.51**| 0.84**| 0.84**| 1.00  | 1.00|
| TC-a        | -0.40**| 0.22**| 0.22** | 0.22** | 1.00    | 0.23**| 0.23**| 0.80**| 0.84**| 0.84**| 1.00  | 1.00|
| TB-Cr       | -0.52**| 0.14  | -0.16* | -0.15  | 0.41**  | 0.32**| 0.28**| 0.28**| 0.71**| 0.70**| 0.58**| 1.00|
| TC-Cr       | -0.53**| 0.00  | -0.02  | -0.01  | 0.47**  | 0.50**| 0.48**| 0.50**| 0.70**| 0.70**| 1.00  | 1.00|
| t4          | 0.11  | 0.17* | -0.18* | 0.18*  | 0.18*   | 0.18*| 0.18*| 0.18*| -0.12 | -0.12 | 0.12  | 0.12|
| t5          | 0.11  | 0.12  | 0.12   | 0.12   | 0.12    | 0.07 | 0.08 | 0.07 | -0.08 | -0.13 | -0.16*| -0.16|

* P < 0.05, ** P < 0.01

The result showed that: (1) There was a strong correlation between the tongue coating texture parameters, the color space parameters of the tongue coating and tongue body were also correlated. The correlation between the tongue coating texture parameters and the color space parameters was weaker than the correlation of the pulse parameters. (2) There was a definite correlation between pulse parameters t4 and tongue parameters TC-ASM, TC-ENT and TC-MEAN, the correlation coefficients were − 0.18, 0.18 and 0.18, respectively. (3) There was a weak correlation between t5 and TB-Cr, and the correlation coefficient was − 0.16 (P < 0.05).

The heat map result of Yin deficiency syndrome was shown in Fig. 5.

The correlation analysis result of tongue and pulse data between Yin deficiency syndrome was shown in Table 5.
Table 5
Correlation analysis of tongue and pulse diagnosis data of Yin deficiency syndrome

| Index | perAll | TC-Con | TC-ASM | TC-ENT | TC-MEAN | TB-S | TC-S | TB-a | TC-a | TB-Cr | TC-Cr | t4 | t5 |
|-------|--------|--------|--------|--------|---------|------|------|------|------|------|-------|-----|-----|
| perAll | 1.00   |        |        |        |         |      |      |      |      |      |       |     |     |
| TC-Con | 0.27** | 1.00   |        |        |         |      |      |      |      |      |       |     |     |
| TC-ASM | -0.34**| -0.95**| 1.00   |        |         |      |      |      |      |      |       |     |     |
| TC-ENT | 0.34** | 0.96** | -1.00**| 1.00   |         |      |      |      |      |      |       |     |     |
| TC-MEAN | 0.27** | 0.98** | -0.97**| 0.97** | 1.00   |      |      |      |      |      |       |     |     |
| TB-S  | -0.59**| -0.40**| 0.45** | -0.45**| -0.41**| 1.00 |      |      |      |      |       |     |     |
| TC-S  | -0.62**| -0.52**| 0.57** | -0.58**| -0.53**| 0.75**| 1.00 |      |      |      |       |     |     |
| TB-a  | -0.64**| -0.33**| 0.38** | -0.39**| -0.33**| 0.70**| 0.73**| 1.00 |      |      |       |     |     |
| TC-a  | -0.62**| -0.48**| 0.50** | -0.51**| -0.46**| 0.67**| 0.83**| 0.82**| 1.00 |      |       |     |     |
| TB-Cr | -0.79**| -0.27**| 0.34** | -0.34**| -0.28**| 0.64**| 0.69**| 0.80**| 0.75**| 1.00 |       |     |     |
| TC-Cr | -0.67**| -0.180*| 0.19* | -0.20**| -0.15 | 0.43**| 0.56**| 0.59**| 0.74**| 0.75**| 1.00 |     |     |
| t4    | 0.03   | 0.10   | -0.14 | 0.13   | -0.10 | -0.09 | -0.10 | -0.14 | -0.08 | -0.08 | 1.00 |     |     |
| t5    | 0.15   | 0.18   | -0.20**| 0.20   | 0.19* | -0.21**| -0.27**| -0.23**| -0.33**| -0.21**| -0.23**| 0.73**| 1.00 |

**P < 0.05, ***P < 0.01

The result showed that: (1) Similar to Qi deficiency syndrome, the tongue coating texture parameters of the Yin deficiency syndrome have a strong correlation, and the color space parameters of the tongue coating and tongue body are also strongly correlated. The correlation between tongue coating texture parameters and color space parameters was weaker than that of pulse parameters. (2) There was a certain correlation between pulse parameters t4 and tongue parameters TC-ASM and TC-a, the correlation coefficients were both -0.14, but the difference was not statistically significant (P > 0.05). (3) t5 was strongly correlated with TB-a, TC-S, TC-Cr and TB-a, and the correlation coefficients were -0.33, -0.27, -0.23, -0.23, respectively (P < 0.01). The correlation coefficients of t5 with TB-Cr, TB-S and TC-ASM were -0.21, -0.20, -0.20, respectively (P < 0.01).

Comparing the results of tongue and pulse correlation analysis between Qi deficiency syndrome and Yin deficiency syndrome, it could be seen that the tongue and pulse correlation intensity of Yin deficiency syndrome was significantly stronger than that of Qi deficiency syndrome, and the correlation between t4 and tongue parameters in Yin deficiency syndrome was significantly reduced, while the correlation between t5 and tongue parameters was significantly increased.

**Machine Learning Results**

Based on Neural Network, Random Forest, SVM, Logistic Regression four machine learning methods, the modeling result of Qi deficiency syndrome and Yin deficiency syndrome based on symptom, tongue & pulse data, symptom & tongue & pulse data was shown in Table 6.
Table 6

| Data sets                  | Model          | AUC  | sensitivity | specificity | F1    | precision | accuracy |
|----------------------------|----------------|------|-------------|-------------|-------|-----------|----------|
| Symptom                    | Neural Network | 0.9223 | 0.9063     | 0.8286     | 0.8657| 0.8286    | 0.8657   |
|                            | SVM            | 0.9321 | 0.8750     | 0.8857     | 0.8750| 0.8750    | 0.8806   |
|                            | Logistic Regression | 0.9000 | 0.8125     | 0.8286     | 0.8125| 0.8125    | 0.8209   |
|                            | Random Forest  | 0.9116 | 0.7813     | 0.8571     | 0.8065| 0.8333    | 0.8209   |
| Tongue & Pulse             | Neural Network | 0.7677 | 0.6316     | 0.6897     | 0.6761| 0.7273    | 0.6567   |
|                            | SVM            | 0.7455 | 0.6842     | 0.6552     | 0.7027| 0.7222    | 0.6716   |
|                            | Logistic Regression | 0.8022 | 0.6842     | 0.8276     | 0.7536| 0.8387    | 0.7463   |
|                            | Random Forest  | 0.7314 | 0.5263     | 0.8621     | 0.6452| 0.8333    | 0.6716   |
| Symptom & Tongue & Pulse   | Neural Network | 0.9401 | 0.9310     | 0.8421     | 0.8710| 0.8182    | 0.8806   |
|                            | SVM            | 0.9328 | 0.6552     | 0.9737     | 0.7755| 0.9500    | 0.8358   |
|                            | Logistic Regression | 0.9301 | 0.7931     | 0.8684     | 0.8070| 0.8214    | 0.8358   |
|                            | Random Forest  | 0.9229 | 0.8966     | 0.8421     | 0.8525| 0.8125    | 0.8657   |

The ROC curves of the models based on symptom, tongue and pulse, and symptom & tongue and pulse were shown in Fig. 6, Fig. 7, and Fig. 8, respectively.

According to the above modeling results, the classification efficiency of each model based on different data sets had the following order: tongue & pulse model < symptom model < symptom & tongue & pulse model. Among them, the SVM model had a better classification performance for symptom data sets, and the area under the ROC curve was 0.9321. Logistic Regression model had a better classification performance for tongue & pulse data sets, with an area under the ROC curve of 0.9401. Neural Network model had a better classification performance for the symptom & tongue & pulse data sets, and with an area under the ROC curve of 0.9401.

Discussion

Treatment based on syndrome differentiation is the basic principle of TCM to recognize and treat diseases, and it runs through the whole process of prevention and rehabilitation of medical care practices. Syndrome differentiation is to recognize the disease and determine the syndrome, treatment is to establish treatment methods and prescription drugs based on the results of syndrome differentiation. Syndrome differentiation is the prerequisite and basis for treatment. Only on the basis of accurate syndrome differentiation can we get a good therapeutic effect. Qi deficiency syndrome and Yin deficiency syndrome are two common syndromes in TCM. According to the basic theory of TCM syndrome differentiation, Qi deficiency syndrome refers to the lack of vitality of the body and the decreased function of visceral organs. The main manifestations are fatigue, lack of energy, lazy speech, and weak pulse. Yin deficiency syndrome refers to the lack of yin fluid in the human body, its nourishing and nourishing functions are reduced, or Yin does not control Yang, Yang is too hyperactive. The main manifestations are dry mouth and pharynx, dysphoria in chest-palms-soles, tidal fever and night sweating. In the classification of lung cancer syndromes, the main manifestations of Qi deficiency syndrome are: cough, white or foamy phlegm, small amount of hemoptysis, chest tightness, shortness of breath, low fever, spontaneous sweating, lack of energy, pale complexion, poor appetite, loose stools, pale red tongue with tooth marks, thin white coating, thin pulse. Yin deficiency syndrome is mainly manifested as: cough without phlegm, or less but sticky phlegm, phlegm with blood, shortness of breath and dull chest pain, low fever, dry mouth, night sweat, upset and insomnia, red tongue, little or bare without tongue coating, thin and rapid pulse. According to the principle of TCM syndrome differentiation and treatment, the principle and treatment method of Qi deficiency syndrome is to invigorate the spleen and replenish qi, and the corresponding prescription is Sijunzi decoction. The principle and treatment method of Yin deficiency syndrome is to nourish Yin and clear lung, and the corresponding prescription is Shashen Maidong decoction.

Statistical Analysis of tongue and pulse data of Qi deficiency and Yin deficiency

TCM is a promising and effective adjuvant therapy in the treatment of lung cancer. Compared with chemotherapy and radiotherapy, it has the advantages of availability, effectiveness and low toxicity\(^33\), its various mechanisms deserve further study\(^{34–35}\). In this study, the tongue parameters TB-a, TC-a, TB-Cr, TC-Cr of Qi deficiency syndrome and Yin deficiency syndrome, thy all represent the red value of tongue body and tongue coating, the larger the value, the more redder or magenta the tongue is. In Yin deficiency syndrome, TB-a, TC-a, TB-Cr, and TC-Cr were
Conclusions

all higher than those in Qi deficiency syndrome, indicating that the tongue of Yin deficiency syndrome was redder or magenta. S stands for saturation, the higher the value of S, the brighter the tongue color will be. TC-S in Yin deficiency syndrome was higher than that in Qi deficiency syndrome, indicating that the tongue color of Yin deficiency syndrome was brighter. perAll is the ratio of tongue coating area to total tongue area. perAll has a higher diagnostic value for thick coating, the higher the value, the thicker the tongue coating. perAll in Yin deficiency syndrome was lower than that in Qi deficiency syndrome, indicating that the tongue coating was thinner in Yin deficiency syndrome. Among the four parameters of texture parameters Con, ASM, ENT, and MEAN, the smaller the value of Con, ENT, and MEAN, the larger the ASM, reflecting that the more delicate the tongue texture or the more greasy the tongue coating. In this study, TC-Con and TC-ENT of Yin Deficiency Syndrome were significantly lower than those of Qi deficiency syndrome, while TC-ASM was higher than that of Qi deficiency syndrome, indicating that the tongue coating of Yin deficiency syndrome was more greasy.

In the pulse parameters, t4 is the time value from the starting point to the descending isthmus of the sphygmogram, corresponding to the systolic period of the left ventricle, and t5 is the time value from the dicrotic notch to the end point of the sphygmogram, corresponding to the diastolic period of the left ventricle. t4 and t5 of Yin deficiency syndrome were smaller than those of Qi deficiency syndrome, indicating that the time of systole and diastole of Yin deficiency syndrome were shorter than those of Qi deficiency syndrome, the pulsation cycle t of Yin deficiency syndrome also showed a decreasing trend, indicating that the pulse wave velocity of Yin deficiency syndrome was slightly higher. In addition, there was a phenomenon of elevation of Yin deficiency syndrome in dicrotic notch h4. In addition, indicrotic notch h4 in Yin deficiency syndrome was elevated. In the Qi deficiency syndrome, h3/h1, h1/t1, and t1 were prolonged, reflecting that the pulse force of the Qi deficiency syndrome was soft and weak, the amplitude of main wave h1 was reduced, and the area under the sphygmogram was smaller, indicating that the pulse body was thin and small. All in all, the tongue of Qi deficiency syndrome was pale and the pulse was weak, while the tongue body of Yin deficiency syndrome was more red or crimson, more brighter in tongue color, thinner and greasy in tongue coating, and more fine in pulse.

Tongue and pulse data modeling analysis of Qi deficiency syndrome and Yin deficiency syndrome

In recent years, with the rapid development of computer technology, different recognition algorithm and machine learning methods, such as Logical Regression[36], SVM[22, 37], Random Forest[38] and neural network[15, 39] and other data mining technologies have been widely used in medical research. The quantitative diagnosis of diagnostic information through various mathematical models has promoted the development of TCM informatization. In this study, symptom and tongue and pulse data were used to classify syndromes. The results showed that the classification efficiency of models based on different data sets was as follows: model of tongue & pulse data < model of symptom < model of symptom & tongue & pulse data, indicating that tongue and pulse data contributed to the classification of syndrome to some extent. Therefore, when faced with a complicated quantitative and qualitative, subjective and objective, determine and fuzzy, massive TCM data combining linear and nonlinear. TCM syndrome associated with complex multidimensional characteristics, and associated with multiple micro index, especially when symptoms were not evident, to explore the relationship between different syndromes and physical and chemical indexes can effectively assist in syndrome differentiation. Research also shows that it is very reasonable to combine micro indexes with macro symptoms. Using machine learning or data mining methods to build TCM syndrome or disease diagnosis model can make the process of syndrome differentiation and treatment more objective, standardized and intelligent[40–42].

The deficiency of this study is that the sample size is insufficient. In future studies, large-scale, multi-center and large-sample size studies will be helpful for further exploration. In addition, this study only analyzed the tongue and pulse characteristics and classification model of Qi deficiency syndrome and Yin deficiency syndrome of NSCLC, in the future, dialectical studies based on more kinds of syndromes of NSCLC are needed.

Limitations And Future Work

This research is based on the real-world investigation, and the results basically conform to the syndrome distribution feature of NSCLC in the clinic. However, there are also some limitations in the study. First of all, due to the limitation of time and place, the sample size of this study was not large enough. Secondly, the basic data statistics of the subjects are not comprehensive enough, and there is a lack of statistics on height, weight, body mass index (BMI), history of present illness, and past medical history, etc., which may affect the data results. Last but not the least, this study mainly focused on the common non-small cell lung cancer syndrome of Qi deficiency and Yin deficiency, lack of more syndromes to explore. In the future, a large-scale and multicenter epidemiological investigation should be combined, the collection of four diagnostic information and basic characteristics needs to be more standardized and complete, and further researches based on more comprehensive syndrome differentiation results need to be carried out.

Conclusions
To sum up, objective tongue and pulse data of NSCLC are useful for the classification of TCM syndrome, they can improve the accuracy of TCM syndrome classification to a certain extent. Tongue and pulse diagnosis parameters can provide new ideas and methods for TCM syndrome differentiation of Qi deficiency syndrome and Yin deficiency syndrome of NSCLC.

**Abbreviations**

NSCLC: non-small cell lung cancer; TCM: Traditional Chinese Medicine; NCCN: national comprehensive cancer network; WHO: World Health Organization; ANOVA: analysis of variance; SVM: support vector machine; TP: true positive; TN: true negative; FP: false positives; FN: false negative; BMI: body mass index.

**Declarations**

**Author's contributions**

YS and JL drafted the initial manuscript and performed classification models, XH, LT and JC performed data statistical analyses, JX and LX developed the study concept and design. JL, ZB, and JL provided critical revisions. All authors read and approved the final manuscript.

**Author details**

1. Basic Medical College, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Pudong, Shanghai, China.
2. Shanghai Innovation Center of TCM Health Service, Shanghai University of Traditional Chinese Medicine, 1200 Cailun Road, Pudong, Shanghai, China.
3. Shanghai University of Traditional Chinese Medicine Yueyang Hospital of Integrated Traditional Chinese Medicine and Western Medicine, 110 Ganhe Road, Hongkou, Shanghai, China.

**Acknowledgments**

We want to thank all the medical staff for their positive support received from the Medical Examination Center of Shuguang Hospital affiliated to Shanghai University of Traditional Chinese Medicine and Shanghai University of Traditional Chinese Medicine Yueyang Hospital of Integrated Traditional Chinese Medicine and Western Medicine.

**Competing interests**

The authors declare that they have no competing interests.

**Availability of data and materials**

The datasets used and/or analyzed during the current study are available from the corresponding author on request.

**Consent for publication**

Not applicable.

**Ethical approval and consent to participate**

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

**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).

**References**

1. Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends—An Update. Cancer epidemiology, biomarkers & prevention: a publication of the American Association for Cancer Research. cosponsored by the American Society of Preventive Oncology. 2016;25(1):16–27.
2. Torre LA, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A. Global cancer statistics, 2012. Cancer J Clin. 2015;65(2):87–108.

3. World Health Organization Cancer. 2020.

4. Liu G, Pei F, Yang F, Li L, Amin AD, Liu S, et al. Role of Autophagy and Apoptosis in Non-Small-Cell Lung Cancer. International journal of molecular sciences 2017;18(2).

5. Travis WD, Brambilla E, Noguchi M, Nicholson AG, Geisinger KR, Yatabe Y, et al. International association for the study of lung cancer/american thoracic society/european respiratory society international multidisciplinary classification of lung adenocarcinoma. Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer. 2011;6(2):444–85.

6. [Chinese expert consensus on antiangiogenic drugs for advanced non-small cell lung cancer. (2020 Edition)]. Zhonghua zhong liu za zhi [Chinese journal of oncology] 2020; 42(12): 1063-77.

7. Miller KD, Nogueira L, Mariotto AB, Rowland JH, Yabroff KR, Alfano CM, et al. Cancer treatment and survivorship statistics, 2019. Cancer J Clin. 2019;69(5):363–85.

8. Choo JR, Soo RA. Lorlatinib for the treatment of ALK-positive metastatic non-small cell lung cancer. Expert Rev Anticancer Ther. 2020;20(4):233–40.

9. Islam KM, Anggondowati T, Deviany PE, Ryan JE, Fetrick A, Bagenda D, et al. Patient preferences of chemotherapy treatment options and tolerance of chemotherapy side effects in advanced stage lung cancer. BMC Cancer. 2019;19(1):835.

10. Chen S, Bao Y, Xu J, Zhang X, He S, Zhang Z, et al. Efficacy and safety of TCM combined with chemotherapy for SCLC: a systematic review and meta-analysis. J Cancer Res Clin Oncol. 2020;146(11):2913–35.

11. Liu R, He SL, Zhao YC, Zheng HG, Li CH, Bao YJ, et al. Chinese herbal decoction based on syndrome differentiation as maintenance therapy in patients with extensive-stage small-cell lung cancer: an exploratory and small prospective cohort study. Evidence-based complementary and alternative medicine: eCAM 2015; 2015: 601067.

12. Chen S, Flower A, Ritchie A, Liu J, Molassiotis A, Yu H, et al. Oral Chinese herbal medicine (CHM) as an adjuvant treatment during chemotherapy for non-small cell lung cancer: A systematic review. Lung cancer (Amsterdam Netherlands). 2010;68(2):137–45.

13. Jiang M, Lu C, Zhang C, Yang J, Tan Y, Lu A, et al. Syndrome differentiation in modern research of traditional Chinese medicine. J Ethnopharmacol. 2012;140(3):634–42.

14. Xia S, Zhang J, Du G, Li S, Vong CT, Yang Z, et al. A Microcosmic Syndrome Differentiation Model for Metabolic Syndrome with Multilabel Learning. Evidence-based complementary and alternative medicine: eCAM 2020; 2020: 9081641.

15. Li X, Zhang Y, Cui Q, Yi X, Zhang Y. Tooth-Marked Tongue Recognition Using Multiple Instance Learning and CNN Features. IEEE transactions on cybernetics. 2019;49(2):380–7.

16. Qin B, Liang L, Wu J, Quan Q, Wang Z, Li D. Automatic Identification of Down Syndrome Using Facial Images with Deep Convolutional Neural Network. Diagnostics (Basel, Switzerland) 2020; 10(7).

17. Pan Z, Shen Z, Zhu H, Bao Y, Liang S, Wang S, et al. Clinical application of an automatic facial recognition system based on deep learning for diagnosis of Turner syndrome. Endocrine 2020.

18. Xia X, Chen X, Wu G, Li F, Wang Y, Chen Y, et al. Three-dimensional facial-image analysis to predict heterogeneity of the human ageing rate and the impact of lifestyle. Nature metabolism. 2020;2(9):946–57.

19. Bing F, Shaozi L. Unsupervised clustering analysis of human-pulse signal in Traditional Chinese Medicine. CAAI Transactions on Intelligent Systems. 2018;13(04):564–70.

20. Wang X, Zhang B, Yang Z, Wang H, Zhang D. Statistical analysis of tongue images for feature extraction and diagnostics. IEEE transactions on image processing: a publication of the IEEE Signal Processing Society. 2013;22(12):5336–47.

21. Kamarudin ND, Ooi CY, Kawanabe T, Odaguchi H, Kobayashi FA, Fast SVM-Based Tongue’s Colour Classication Aided by k-Means Clustering Identiers and Colour Attributes as Computer-Assisted Tool for Tongue Diagnosis. Journal of healthcare engineering 2017; 2017: 7460168.

22. Zhang J, Xu J, Hu X, Chen Q, Tu L, Huang J, et al. Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images. BioMed research international 2017; 2017: 7961494.

23. Wood DE. National Comprehensive Cancer Network (NCCN) Clinical Practice Guidelines for Lung Cancer Screening. Torac Surg Clin. 2015;25(2):185–97.

24. Brambilla E, Travis WD, Colby TV, Corrin B, Shimosato Y. The new World Health Organization classification of lung tumours. Eur Respir J. 2001;18(6):1059–68.

25. Micke P, Mattsson JS, Djureinovic D, Nodin B, Jirström K, Tran L, et al. The Impact of the Fourth Edition of the WHO Classification of Lung Tumours on Histological Classification of Resected Pulmonary NSCCs. Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer. 2016;11(6):862–72.
26. Bei H. Technical Guidelines for Clinical Research of New Drugs of Syndrome. J Tradit Chin Med. 2018;26(21):107.
27. Institute of TCM Diagnosis HUoT, Department of Internal Medicine CAoT, Department of Surgery CAoT, et al. Clinic terminology of traditional Chinese medical diagnosis and treatment—Syndromes. State Bureau of Technical Supervision; 1997. p. 92.
28. Jian-feng Z, Jia-tuo X, Li-ping T, Zhi-feng Z, Xiao-juan H, Ji C, et al. Study on the Characteristics of Sub-health Symptoms and TCM Syndrome Patterns Distribution in 1 754 Non-disease Population Chinese. Journal of Integrative Medicine. 2017;37(08):934–8.
29. Schiller F, Valsecchi M, Gegenfurtner KR. An evaluation of different measures of color saturation. Vision research. 2018;151:117–34.
30. Sun X, Young J, Liu JH, Bachmeier L, Somers RM, Chen KJ, et al. Prediction of pork color attributes using computer vision system. Meat Sci. 2016;113:62–4.
31. Qi Z, Tu LP, Chen JB, Hu XJ, Xu JT, Zhang ZF. The Classification of Tongue Colors with Standardized Acquisition and ICC Profile Correction in Traditional Chinese Medicine. BioMed research international 2016; 2016: 3510807.
32. Luo ZY, Cui J, Hu XJ, Tu LP, Liu HD, Jiao W, et al. A Study of Machine-Learning Classifiers for Hypertension Based on Radial Pulse Wave. BioMed research international 2018; 2018: 2964816.
33. Xiang Y, Guo Z, Zhu P, Chen J, Huang Y. Traditional Chinese medicine as a cancer treatment: Modern perspectives of ancient but advanced science. Cancer medicine. 2019;8(5):1958–75.
34. Su XL, Wang JW, Che H, Wang CF, Jiang H, Lei X, et al. Clinical application and mechanism of traditional Chinese medicine in treatment of lung cancer. Chin Med J. 2020;133(24):2987–97.
35. Leng J, Lei L, Lei SF, Zhu Z, Ocampo A, Gany F. Use of Traditional Chinese Herbal Medicine Concurrently with Conventional Cancer Treatment Among Chinese Cancer Patients. Journal of immigrant minority health. 2020;22(6):1240–7.
36. Zhang K, Geng W, Zhang S. Network-based logistic regression integration method for biomarker identification. BMC systems biology. 2018;12(Suppl 9):135.
37. Liu C, Cheng Y. An Application of the Support Vector Machine for Attribute-By-Attribute Classification in Cognitive Diagnosis. Applied psychological measurement. 2018;42(1):58–72.
38. Kong Y, Yu T. A Deep Neural Network Model using Random Forest to Extract Feature Representation for Gene Expression Data Classification. Scientific reports. 2018;8(1):16477.
39. Yu L, Chen H, Dou Q, Qin J, Heng PA. Automated Melanoma Recognition in Dermoscopy Images via Very Deep Residual Networks. IEEE Trans Med Imaging. 2017;36(4):994–1004.
40. Xia S, Cai J, Chen J, Lin X, Chen S, Gao B, et al. Factor and Cluster Analysis for TCM Syndromes of Real-World Metabolic Syndrome at Different Age Stage. Evidence-based complementary and alternative medicine: eCAM 2020; 2020: 7854325.
41. Gu Y, Wang Y, Ji C, Fan P, He Z, Wang T, et al. Syndrome Differentiation of IgA Nephropathy Based on Clinicopathological Parameters: A Decision Tree Model. Evidence-based complementary and alternative medicine: eCAM 2017; 2017: 2697560.
42. Kang H, Zhao Y, Li C, Chen Y, Tang K, Yang L, et al. Integrating clinical indexes into four-diagnostic information contributes to the Traditional Chinese Medicine (TCM) syndrome diagnosis of chronic hepatitis B. Scientific reports. 2015;5:9395.

**Figures**
Figure 1
Research flowchart
Figure 2

The TFDA-1 digital tongue diagnosis instrument and its corresponding tongue image analysis system

Figure 3

PDA-1 digital pulse diagnosis instrument and its corresponding sphygmogram
Figure 4
Heat map of tongue and pulse correlation analysis of Qi deficiency syndrome
Figure 5

Heat map of tongue and pulse correlation analysis of Yin deficiency syndrome
Figure 6

ROC curves of Qi deficiency syndrome model based on symptom

Figure 7

ROC curves of Qi deficiency syndrome model based on tongue & pulse data
Figure 8

ROC curves of Qi deficiency syndrome model based on syndrome & tongue & pulse data