Clinical data mining on network of symptom and index and correlation of tongue-pulse data in fatigue population

yulin Shi
Shanghai University of Traditional Chinese Medicine https://orcid.org/0000-0003-4379-8561

Xiaojuan Hu
Shanghai University of Traditional Chinese Medicine

Cui Ji
Shanghai University of Traditional Chinese Medicine

Longtao Cui
Shanghai University of Traditional Chinese Medicine

Jingbin Huang
Shanghai University of Traditional Chinese Medicine

Xuxiang Ma
Shanghai University of Traditional Chinese Medicine

Jiang Tao
Shanghai University of Traditional Chinese Medicine

Xinghua Yao
Shanghai University of Traditional Chinese Medicine

Jiatuo Xu (xjt@fudan.edu.cn)
https://orcid.org/0000-0002-3498-2132

Fang Lan
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

Yu Wang
Shanghai University of Traditional Chinese Medicine

Hongyuan Fu
Shanghai University of Traditional Chinese Medicine

Jue Wang

Shanghai University of Traditional Chinese Medicine
Yanting Lin
Shanghai University of Traditional Chinese Medicine
Jingxuan Bai
Shanghai University of Traditional Chinese Medicine
Xiaojing Guo
Shanghai University of Traditional Chinese Medicine
Liping Tu
Shanghai University of Traditional Chinese Medicine

Research article

Keywords: Fatigue, Complex network, Symptom, Index, Tongue and pulse data

Posted Date: January 7th, 2021

DOI: https://doi.org/10.21203/rs.3.rs-59498/v2

Version of Record: A version of this preprint was published on February 24th, 2021. See the published version at https://doi.org/10.1186/s12911-021-01410-3.

Abstract

Background: Fatigue is a kind of non-specific symptom, which occurs widely in sub-health and various diseases. It is closely related to people's physical and mental health. Due to the lack of objective diagnosis criteria, it is often neglected in clinical diagnosis, especially in the early disease stage. Many clinical practices and research have shown that tongue and pulse conditions reflect the body's overall state. Establishing an objective evaluation method for diagnosing disease fatigue and non-disease fatigue by combining clinical symptoms, indexes, and tongue & pulse data is of great significance for timely and effective clinical treatment.

Methods: In this study, 2632 physical examination populations were divided into healthy controls, sub-health fatigue group, and disease fatigue group. Complex network technology was used to screen out the core symptoms and Western medicine indexes of sub-health fatigue and disease fatigue populations. Pajek software was used to construct the core symptoms/indexes network and core symptoms-indexes combined network. Simultaneously, the canonical correlation analysis method was used to analyze the objective tongue & pulse data between the two groups of fatigue population and analyze the distribution of tongue & pulse data.

Results: Some similarities were found in the core symptoms of sub-health fatigue and disease fatigue population, but with different node importance. The node-importance difference indicated that the diagnostic contribution rate of the same symptom to the two groups was different. The canonical correlation coefficient of tongue & pulse data in the disease fatigue group was 0.42 ($P < 0.05$). On the contrast, correlation analysis of tongue & pulse in the sub-health fatigue group showed no statistical significance.

Conclusions: The complex network technology was suitable for the correlation analysis of symptoms and indexes in the fatigue population, and the tongue & pulse data had a certain diagnostic contribution to the classification of the fatigue population.

Name of the registry: Chinese Clinical Trial Registry

Trial registration number: ChiCTR-IOR-15006502; ChiCTR1900026008

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)

Background

Fatigue is a state in which the body cannot initiate or maintain a certain intensity of activity, or manifest a pathological dysfunction during the initiation or maintenance of voluntary activity. It is a physiological

manifestation of the body's self-regulation, and a pathological result of certain diseases[1]. Fatigue is a non-specific symptom with certain heritability[2]. It is ubiquitous in sub-health and various diseases, such as Parkinson's[3], major depressive disorder[4], schizophrenia[5], and cancer[6]. It has become one of the main factors harmful to human's physical and mental health with serious decline in work efficiency and life quality.

Sub-health status is defined as the decline in vitality, physiological function, and adaptation capacity, but in general not meeting the diagnostic criteria for clinical or sub-clinical disease[7; 8]. It has attracted extensive attention with the increasing incidence of sub-health in recent years. Sub-health fatigue is one of the most common sub-health types with that fatigue as the chief complaint. The etiology and pathogenesis of fatigue are largely unknown. There is still a lack of objective, effective, and comprehensive evaluation methods for diagnosing fatigue, leading to the inability to carry out targeted interventions when fatigue occurs in the early stage of diseases. Hence it is imperative to establish a comprehensive, objective evaluation method for fatigue.

In the diagnosis methods of Traditional Chinese Medicine (TCM), tongue & pulse diagnosis always plays an important role in clinical diagnosis and treatment. Tongue & pulse diagnosis are comprehensive diagnostic methods based on the body's overall state, suitable for the comprehensive evaluation of the body's functional state, and have become an important objective basis for health status evaluation and syndrome diagnosis[9]. However, traditional tongue & pulse diagnosis lack data classification accuracy. on the one hand, artificial intelligence and machine learning methods can quickly and accurately complete the basic data cleaning and large-scale data sorting, and on the other hand, do data mining for massive tongue & pulse data. Different recognition algorithms and machine learning methods have been widely used in image recognition, target detection, natural language processing, and other fields[10-12]. Nowadays, breakthroughs have been made in fatigue quantification and standardization. Artificial Neural Network[13], Support Vector Machine[14], K Nearest Neighbor[15], and other machine learning methods have helped to achieve the digitalization of TCM tongue & pulse diagnosis and the establishment of the corresponding disease diagnosis model[16; 17]. The diagnostic relationship between tongue & pulse and health state can be better established through accurate detection, identification, and multi-dimensional quantitative analysis of tongue & pulse data to save medical resources and improve diagnosis efficiency and treatment efficacy[18-20]. Data-driven researches on fatigue diagnosis technology using tongue & pulse data have been increasing day by day. Researches based on tongue[21; 22] and pulse[23-25] of fatigue population have shown that their tongue and pulse had their own unique characteristics.

Complex network is a basic framework with high topological abstraction. Further analysis of the network through classification, screening, and other analytical methods can mine the potential rules of many clinical data. Complex networks are used in the analysis of basic rules of TCM[26], network pharmacology[27-29], combined analysis of TCM syndromes and network pharmacology[30; 31], and symptom evolution[32]. Complex networks are mostly used to construct qualitative network relations, which are more suitable for analyzing the symptoms involved in identifying certain syndromes and the relationship between symptoms. Fatigue is a complex concept, and it is one of the symptoms that can

most reflect the interaction between psychology and physiology. As a non-specific symptom, different network relationships of fatigue symptom can be found in depression network[33], qi-deficiency syndrome of coronary heart disease network [34; 35], and qi-deficiency syndrome of breast cancer network[36].

Despite the progress made in current research on fatigue, there are still many problems that need to be solved. For example, the studies on fatigue are relatively simple. It is essential to study the correlation of fatigue symptoms and do the combined analysis of fatigue symptoms and indexes. Besides, the analysis of objective tongue & pulse data of fatigue is mostly independent. To address the above problems, in this study, 2362 physical examination populations were divided into healthy controls, sub-health fatigue group, and disease fatigue group. Complex network technology was used to screen out the core symptoms and Western medicine indexes. Through constructing the core symptoms and indexes network and the core symptoms-indexes combined network, analyzing the network structure to establish the distribution of fatigue symptom and index. Simultaneously, the canonical correlation analysis method was used to get the associated relationship between tongue and pulse data of disease fatigue and sub-health fatigue population. Based on symptoms, Western medicine indexes, and tongue & pulse data, this study tried to explore the different fatigue populations' characteristics from different dimensions.

Methods

Study design

All the 7,025 individuals were selected in the Medical Examination Center of Shuguang Hospital affiliated to Shanghai University of Traditional Chinese Medicine from Jul. 2015 to Dec.2018. A total of 361 in the sub-health fatigue population and 1529 in the disease fatigue population were further selected.

Four well-trained clinicians diagnosed the disease according to the diagnostic criteria of western medicine. The Health Status Assessment Questionnaire (shortly, H20) and Information Record Form of Four Diagnosis of Traditional Chinese Medicine (Copyright No.: 2016Z11L025702) designed by the sub-health research group of the "863 Plan" was used to investigate fatigue syndrome. Patients with positive fatigue symptom were defined as disease fatigue population, those who had no obvious positive indexes in Western medicine, H20 score was between 60-79, and with positive fatigue symptom were defined as sub-health fatigue population, and those who had no obvious positive indexes in Western medicine, H20 score was between 80-100, and without positive fatigue symptom were defined as healthy controls.

The overall flow diagram of this study was shown in Fig. 1.

Information on the Four Diagnostic Methods of TCM and Physical and Chemical Indexes

The four-diagnostic scale of TCM included 25 categories and 256 subitems. The symptoms and signs were classified as none, mild, and severe with 0, 1, and 2 points. We used TFDA-1 (Tongue Face Diagnosis Analysis-1) tongue diagnostic instrument (software copyright registration No.: 2018SR033451)

and PDA-1 (Pulse Diagnosis Analysis-1) pulse diagnostic instrument (Patent No.: ZL201620157027.6), which independently developed by the National Key Research and Development Program "Traditional Chinese Medicine Intelligent Tongue Diagnosis System Research and Development" to collect the clinical tongue and pulse data. Besides, clinical Western medicine indexes mainly include 191 items such as blood routine, urine routine, liver and kidney function, tumor markers, electrocardiogram, and imaging examination. Considering fatigue had no specific indexes, the existing western medicine indexes were auxiliary. It was necessary to combine specific symptoms in specific diseases for comprehensive analysis and judgment. For example, anemia fatigue patients were often associated with decreased hemoglobin, cancer fatigue patients were often associated with abnormal tumor markers, and hepatopathy fatigue patients were often associated with abnormal liver function indexes. Therefore, in this study, we did not list or describe Western medicine indexes one by one. The tongue diagnosis equipment and the analysis interface were shown in Fig.2 and Fig.3, Fig.2 (A) and Fig.2 (B) were the front and profile view of the tongue diagnosis instrument. The pulse diagnosis instrument and sphygmogram were shown in Fig.4, Fig.4(A) was the PDA-1 pulse diagnosis instrument and supporting equipment, Fig.4(B) was the sphygmogram of PDA-1 pulse diagnosis instrument.

The color parameters of tongue image in Fig. 3 came from four color spaces:[37-39] RGB, HSI, Lab, and YCrCb; they were R (red), G (green) and B (blue), H (hue), S (saturation), I (brightness), L (lightness), a (red-green axis), b (yellow-blue axis), Y (brightness), Cr (the difference between the red part of the RGB input signal and the brightness value of the RGB signal), Cb (the difference between the blue part of RGB input signal and the brightness value of RGB signal), perAll (the ratio of the coating area to the total tongue area) and perPart (the ratio of the coating area to the non-coating tongue area).

The parameters of sphygmogram in figure 4(B), h represented amplitude height, h_1 was the main amplitude, h_3 was the heavy wave front wave amplitude, h_4 was the dicrotic notch amplitude, h_5 was the gravity wave amplitude, t_1 was the time value from the start point to the crest point of the main wave, t_4 was the time value from the start point to the dicrotic notch, t_5 was the time value from the dicrotic notch to the endpoint, t was one pulsating period, and w was divided into w_1 and w_2, w_1 was 1/3 height of the main wave, w_2 was 1/5 height of the main wave.

Normalized Data Entry and Extraction

Using Python3.7 to make data arrangement, and established the symptom and Western medicine index data sets, respectively. A total of 494 symptoms and indexes were collected, including 254 symptoms of TCM and 240 Western medicine indexes. The data set was binarized. The positive TCM syndrome was recorded as "1", and the negative TCM syndrome was recorded as "0". Negative qualitative data of Western medicine index was recorded as "0", positive data including weak positive (+) and strong positive (++ or +++) was recorded as "1", quantitative data in the normal range was recorded as "0", higher than or lower than the normal range was recorded as "1".

Screening of Symptoms and Indexes

According to the data characteristics in this study, the improved node contraction method was used to analyze the network nodes quantitatively. Node contraction was the integration of k nodes connected to the node with this node, replace the k+1 nodes with a new node, and the edges that were previously associated with k+1 nodes were now associated with the new node. After the nodes with high importance were shrunk and fused, the whole network's connection would be closer, and the aggregation degree would increase. This method's basic idea was to shrink the nodes in the network one by one and then compare the network aggregation degree changes to rank the importance of nodes. The improved node contraction method[40] comprehensively considered the weight of edges in the weighted network. In this study, nodes represented symptoms or indexes, when two abnormal symptoms or indexes appeared in the same individual simultaneously, a connection was established between the two nodes. The weight of the edge represented the number of simultaneous occurrences of the two connected nodes. The larger the weight was, the closer the relationship between the two indexes. So here, we choose the sum of the edge weights as the point weights. In the weighted network, the corresponding concept of node degree was node strength. With the specific definition of node strength, extend the definition of network cohesion to weighted networks. Network cohesion refers to the reciprocal of the product of the number of nodes and the average shortest distance. Quantitatively describing the degree of network cohesion, weighted network cohesion[41] was defined as in formula (1).

$$\partial(\mathrm{WG}) = \frac{1}{s \times l} = \frac{1}{\sum_1^n N_i \sum_{j \in N_i} W_{ij} \times \frac{\sum_{i \neq j \in V} d_{ij}}{n \times (n-1)}} \qquad (1)$$

In the above formula, was the sum of the network's average node strength, which was the strength of each node divided by the number of neighboring nodes of the node. was the average shortest distance of the unweighted network corresponding to the weighted network after thresholding. d_{ij} was the shortest distance between node V_i and node V_j in the network. W_{ij} was the weight between node V_i and node V_j. The node importance was expressed by IMC(V_i), which see formula (2), ∂(WG∗V*i*) was the aggregation degree of the weighted network after contraction of node V_i.

$$\mathrm{IMC}(V_i) = 1 - \frac{\partial(WG)}{\partial(WG \times Vi)} \qquad (2)$$

In this study, the improved node contraction method took the degree, betweenness, and edge weight of nodes, which was basically consistent with the purpose and requirement of each node's importance. In this paper, the sum of edge weights was used as the point weights to obtain an undirected weighted network. Triple was an improved form of adjacency list, and many exposed network data was represented in the form of triples. A triple could be thought of as a three-column matrix, with each row in the format of [V_i, V_j, W_{ij}], indicating the existence of an edge with a weight of W from node V_i to node V_j. To better demonstrate the interaction between nodes, the symptom and index pairs were divided by the triad's maximum weight to obtain the normalized weight value.

Construction and Analysis of Complex Network

MATLAB (R2016a) software was used to process the binary data, and the core symptoms and index data were selected according to the importance of nodes. Selected the top 10 data to construct the network and do an analysis by Pajek (Pajek 64 5.08) software. The network diagram was output after editing each node's color and label and adjusting the node position manually.

Statistical Analysis

SPSS (Version 25.0) software was used for statistical processing and analysis. Continuous data with normal distribution were presented as the mean and standard deviation, and those with abnormal distribution were presented as a median and interquartile range. The categorical variables were expressed as counts and percentages. ANOVA was performed for data that normal distribution and homogeneity of variance among groups, Kruskal-Wallis H test was performed for nonnormal distribution data, all the mentioned tests were two-tailed, and a P value < 0.05 was considered statistically significant.

Quality Control

In this study, researchers from Shanghai University of Traditional Chinese Medicine completed all scales and tongue and pulse collection. All researchers were medical professionals in Traditional Chinese Medicine or integrated Traditional Chinese and Western medicine. Moreover, all researchers have been trained in standard operating procedures to ensure consistency and accuracy in interpreting data collection results. Each study participant was interviewed by at least two professional researchers and supervised by at least two senior physicians to ensure data collection consistency and authenticity and reduce measurement bias.

Results

Data Set for Fatigue Group

There were 742 people in the healthy controls, 361 people in the sub-health fatigue group, and 1529 people in the disease fatigue group. The most common diseases in fatigue populations mainly include hypertension, diabetes, hyperlipidemia, and fatty liver, and this randomized trial showed the same pattern. Other diseases, such as coronary heart disease and cancer, can also result in fatigue, but in the Health Examination Center, due to the lack of patients, coronary heart disease and cancer patients were not included. All the subjects in the study were randomly included according to the disease's diagnostic criteria.

The patients in the disease fatigue group were mainly hypertension, diabetes, hyperlipidemia, and fatty liver. The basic statistical characteristics of the healthy controls, sub-health, and disease groups were shown in Table 1. "N" represented the number of categorical variables, "X ± SD" represented the mean and standard deviation of the continuous data Age and BMI.

Table 1. Statistical table of basic information of the participants in the healthy controls, sub-health fatigue, and disease fatigue

Group		N	male	female	Age	BMI
			N（%）	N（%）	(X ± SD, year)	(Kg/m²)
Healthy controls		742	553（74.5）	189（25.5）	32.52±10.16	22.71±3.08
Sub-health fatigue group		361	215（59.6）	146（40.4）	34.64±9.45**	22.74±3.46
Disease fatigue group	Hypertension	311	228(73)	83(27)	48.56±13.94**##	25.51±3.41**##
	Diabetes	157	127(81)	30(19)	54.04±12.79**##	25.95±3.67**##
	Hyperlipemia	518	373(72)	145(28)	45.87±12.69**##	24.97±3.27**##
	Fatty liver	442	334(76)	108(24)	45.10±13.20**##	26.57±3.06**##

*vs. healthy controls, $P < 0.05$, ** vs. healthy controls, $P < 0.01$

#vs. the group of sub-health, $P < 0.05$, ## vs. the group of sub-health, $P < 0.01$

From the result, we could see that there were many more males than females in the study in the three groups, actually the total number of people who participated in medical examinations, males were higher than females, the reason might be that there were more males than females in routine physical examinations in the area where the hospital was located, and men might pay more attention to routine physical examinations. Age and BMI were statistically significant in the Sub-health fatigue group and disease fatigue group subjects compared with Healthy controls (P<0.01), and age was statistically significant in the disease fatigue group compared with the Sub-health fatigue group (P<0.01). Age and BMI, which are closely related to disease, are often mismatched between groups. This study mainly focused on the network relationship of symptom and index and the correlation of tongue and pulse data. This study's objective was independent of age and BMI, so a mismatched age and BMI between the two groups did not affect this study's conclusions.

Construct and Analyze of Symptom Network of the Sub-health Fatigue Group

Using MATLAB for data processing of sub-health fatigue group with the whole symptoms. The binary data of TCM symptom was converted into ". NET" format and then using Pajek software to draw the network, and the symptom network was shown in Fig. 5.

As the total network had many nodes and complex network relationships, its core nodes' relationships could not be well described. Therefore, selected the core nodes according to the importance of nodes IMC(Vi) was necessary, and then used Pajek software to draw the core symptom network. The network was shown in Fig. 6. In the network, the node's size represented the strength, the thickness of the edge represented the weight between nodes, and the core symptoms were shown in table 2.

Table 2 Core symptoms of sub-health fatigue group

Index	Symptom	IMC(Vi)
TC1	white tongue coating	0.999
LP1	headache	0.997
TC2	yellow tongue coating	0.996
QP1	sour	0.996
EM7	dreaminess	0.994
EM3	irritability	0.994
THA4	chest distress	0.992
HE13	xerophthalmia	0.991
TC6	thick coating	0.991
EM6	insomnia	0.989

Analyzed the network, took the symptom associated pairs whose normalized weight was greater than 0.5, and the associated results were shown in table 3.

Table 3 Associated analysis of core symptoms of sub-health fatigue group

Symptom		Weight
LP1	QP1	1.000
TC1	QP1	0.905
	HE13	0.833
	LP1	0.810
	TC6	0.750
	EM3	0.679
	EM6	0.643
	EM7	0.631
TC2	TC6	0.512
QP1	HE13	0.464

Construct and Analyze of Symptom and Index Networks of the Disease Fatigue Group

Using the same method to draw a network of the whole symptom and index of disease fatigue population, as shown in Fig.7. Selected the core symptoms and indexes and draw networks of disease fatigue group. The core symptoms and index networks were shown in Fig. 8 and Fig. 9. The core symptoms and indexes and node importance rank were shown in table 4 and table 5.

Table 4 Core symptom and node importance rank of the disease fatigue group

Index	Symptom	IMC(Vi)
TC1	white tongue coating	1.000
HE1	dizziness	1.000
TC2	yellow tongue coating	1.000
QP1	sour	1.000
EM7	dreaminess	1.000
TC6	thick coating	1.000
PU15	wiry pulse	1.000
TC11	greasy coating	1.000
EM6	insomnia	1.000
EM3	irritability	0.999

Table 5 Core index and node importance rank of the disease fatigue group

Index	Index	IMC□Vi□
BRT13	basophil	1.000
BRT20	platelet distribution width	1.000
SBP	systolic blood pressure	1.000
BRT12	percentage of monocyte	1.000
DBP	diastolic blood pressure	1.000
RUT5	PH of urine	1.000
BRT8	hemoglobin	1.000
BRT10	hematocrit	1.000
BI15	uric acid	1.000
BMI	body mass index	1.000

The relationships between symptoms and indexes were very complicated in the actual clinical diagnosis of disease. The diagnosis could not rely on symptoms or indexes solely. It was necessary to combine them to analyze together. Its core symptom-index interaction edges were shown as cyan lines. The core symptom-index network was shown in Fig.10.

Analyzed the network, took the top 10 pairs of core symptom-symptom pairs and index-index pairs, respectively, and the associated analysis results were shown in Table 6 and Table 7.

Table 6 Associated analysis of core symptom-symptom of the disease fatigue group

Symptom		Weight
TC1	QP1	0.187
TC2	TC6	0.167
TC1	EM3	0.153
TC2	TC11	0.149
TC1	HE1	0.142
TC1	EM7	0.137
PU15	TC1	0.134
TC1	EM6	0.132
TC6	TC11	0.129
TC2	QP1	0.123

Table 7 Associated analysis of core index-index of the disease fatigue group

Index		Weight
BRT13	SBP	1.000
BRT13	BRT20	0.921
BRT13	BRT12	0.913
BRT20	SBP	0.908
SBP	BRT12	0.902
DBP	SBP	0.873
BRT13	DBP	0.868
BRT20	BRT12	0.824
DBP	BRT12	0.782
BRT20	DBP	0.782

Selected the core symptom - index associated results, the top 10 symptom-index pairs were shown in Table 8.

Table 8 Associated analysis of core symptom-index of the disease fatigue group

Symptom	Index	Weight
TC1	SBP	0.546
	BRT13	0.545
	BRT20	0.494
	BRT12	0.491
	DBP	0.474
TC2	BRT13	0.376
	SBP	0.372
	BRT12	0.342
	BRT20	0.338
	DBP	0.326

In conclusion, the research results showed that white coating, yellow coating, sour, dreaminess, irritability, thick coating, and insomnia were the common symptoms of the two groups of fatigue population. The main difference was that the node importance of the same symptom was different in different fatigue population networks, indicating that the same symptom's diagnostic contribution rate to the two groups' population differed. Headache, chest distress, and xerophthalmia were more significant in the sub-health fatigue group, while dizziness, wiry pulse, and greasy coating were more significant in the disease fatigue group. The most common abnormal indexes in the disease fatigue group were basophil, platelet distribution width, systolic blood pressure, percentage of monocyte, diastolic blood pressure, PH of urine, hemoglobin, hematocrit, uric acid, and body mass indicator. The symptom-index associated analysis showed that systolic blood pressure, basophil, platelet distribution width, percentage of monocyte, and diastolic blood pressure were closely related to white coating, and it was also related to yellow coating to some extent.

Canonical Correlation Analysis of Tongue and Pulse Parameters

Individuals with outliers and extreme values in tongue and pulse data of the three groups were excluded. Using SPSS (Version 25.0) software to detect outliers or extreme values. When the distance between the data point and the edge of the box was more than 1.5 times the box body length, it was defined as outliers; when the distance was more than 3 times, it was defined as extreme values. The SPSS system automatically coded the numerical values near outliers or extreme values. Outliers or extreme values affected the research results. In the study, the sample would be deleted once there was an indicator anomaly. Although this reduced the sample size to some extent, it ensured the accuracy of the results. Finally, 551 were included in the healthy controls, 252 in the sub-health fatigue group, and 1,160 in the disease fatigue group.

Canonical Correlation Analysis verified the overall correlation between one set of variables and another. The results showed a certain correlation between the tongue and pulse data in healthy controls and the disease fatigue group. The correlation coefficient of tongue and pulse data in the healthy controls was 0.475 ($P<0.05$), and tongue characteristic parameters were mainly affected by TB-Cb, TB-b, TB-H, and TC-Cb (canonical correlation coefficients were -0.435, 0.431, 0.429, and -0.374, respectively, $P<0.05$). Pulse

characteristic parameters were mainly affected by h_1 and h_1/t_1 (canonical correlation coefficients were 0.388 and 0.378, respectively, $P<0.05$, as shown in Figure 11(A)). The correlation coefficient of tongue and pulse data in the disease fatigue group was 0.420 ($P<0.05$), and tongue characteristic parameters were mainly affected by perAll, TC-Cr, TB-Cr, TB-Cb, TB-b (canonical correlation coefficients were -0.723, 0.697, 0.649, -0.603 and 0.590, respectively, $P < 0.05$). Pulse characteristic parameters were mainly affected by h_4, h_4/h_1, h_3/h_1, h_3 and w_2/t (canonical correlation coefficients were -0.621, -0.609, -0.507, -0.480 and -0.446, respectively, $P < 0.05$, as shown in Fig.11(B). There was no statistically significant correlation between the tongue and pulse data in the sub-health fatigue group.

Discussion

Fatigue is an important early warning signal of abnormal health status. It should be treated in time to prevent its further development into a more serious disease. At present, there are two main reasons for the difficulties in fatigue research. First, the mechanism of fatigue is complex, and there is still a lack of diagnostic criteria for fatigue. Second, there is a lack of effective fatigue evaluation models[42]. Modern tongue and pulse diagnosis can provide good data support, coupled with the assistance of artificial intelligence and technical machine learning technology, providing new methods and ideas for accurate diagnosis of fatigue.

In this study, complex network technology was used to screen out the main symptoms and indexes of fatigue patients in the physical examination population and the interaction between symptoms and indexes. Studying the interrelationship between fatigue-related symptoms was helpful for furtherly determining the diagnosis direction of fatigue-related diseases. Headache, chest distress, and xerophthalmia were more significant in sub-health fatigue. Headache and chest distress were generally manifested as qi stagnation syndrome, and xerophthalmia was the common clinical manifestation of jinye deficient syndrome. Wiry pulse, greasy coating, and dizziness were more significant in the group of disease fatigue. The clinical significance of wiry pulse is mainly about liver and gallbladder disease, pain, phlegm and retained fluid, consumptive disease, and stomach gas decline. The clinical significance of greasy coating was phlegm-damp, phlegm and retained fluid, and dyspepsia. These two symptoms were consistent with common pathological manifestations of the disease. Furthermore, dizziness was the concomitant symptom of hypertension, hypoglycemia, anemia, and cancer. Basophil, platelet distribution width, percentage of monocyte, hemoglobin, hematocrit were blood routine item, the value of PH of urine and uric acid are routine items of urine examination, abnormal of these two indexes mostly indicates the abnormal renal function, and BMI mostly reflects human metabolism, so it can be seen that abnormal blood routine, renal function, blood pressure, and basic metabolism are more common in patients with fatigue.

Associated analysis of symptoms and indexes can better explore the nature of diseases[43; 44]. In this study, the symptom-index associated analysis showed that systolic blood pressure, basophil, platelet distribution width, percentage of monocyte, and diastolic blood pressure were closely related to white coating and yellow coating to some extent. The clinical significance of white coating in disease state was

mainly surface syndrome, cold syndrome, and dampness syndrome. Thus, the fatigue population is mostly seen in surface syndrome, cold syndrome, and dampness syndrome. This analysis was helpful for better understanding the core symptoms, the interaction between symptoms, and the distribution of syndromes of the fatigue groups to provide a theoretical basis for the rapid and accurate diagnosis.

Fatigue as a comprehensive performance of the whole body, it was necessary to analyze the indexes' relationship. The canonical correlation analysis method was used for the combined analysis of tongue and pulse data. Canonical correlation analysis of tongue and pulse data showed that the healthy controls' correlation was stronger than that in the disease fatigue group. In contrast, the correlation coefficient between canonical variables and all tongue and pulse variables in the disease fatigue group was higher than that in the healthy controls. The reason might be that there were many kinds of diseases in fatigue patients, while the healthy controls were relatively single. The tongue and pulse in the healthy controls tended to be more stable, and its characteristics were relatively stable. For example, the healthy controls' tongue was generally reddish tongue, thin white tongue coating, and the pulse was usually normal pulse. The tongue and pulse in the disease fatigue group might be diversified due to different diseases. Patients' tongue could present as purple-red tongue, bluish-purple tongue, yellow greasy tongue coating, white greasy tongue coating, etc. Its pulse could be different forms of wiry pulse, tight pulse, slippery pulse, uneven pulse, etc. The disease fatigue group's tongue and pulse abnormalities destroyed a certain stable correlation of the health state and tended to a certain partial correlation.

This study still has some limitations. Firstly, there are many kinds of diseases in this study, which was not conducive to interpreting the results. In the future, specific diseases can be refined to analyze the symptom-index network's relationship and the tongue-pulse data. Secondly, complexion spectral data can be added based on tongue and pulse data. Integrating more objective indexes that can objectively evaluate fatigue will be more productive to analyze this phenomenon and its mechanism. Besides, this study still lacks treatment guidance and intervention for the fatigue population, which will be improved in the future.

Conclusion

In summary, this study constructed the fatigue-related symptom network and symptom-index network, analyzed the data characteristics of tongue and pulse in fatigue population, the distribution of symptoms, indexes, and revealed the tongue and pulse data in different fatigue populations. It provided an objective basis for establishing the data evaluation of fatigue state, and we are looking forward to establishing a fatigue evaluation method based on objective data of tongue and pulse in the future.

Abbreviations

TCM: Traditional Chinese Medicine; PDA-1: Pulse Diagnosis Analysis-1; TFDA-1: Tongue Face Diagnosis Analysis-1; ANOVA: Analysis of Variance; BMI: Body Mass Index; TB: Tongue Body; TC: Tongue Coating; ALT: Alanine Transaminase; CFS: Chronic Fatigue Syndrome

Declarations

Ethics approval and consent to participate

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.

Consent for publication

Not applicable.

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.

Competing interests

The authors declare that they have no competing interests.

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.

Authors' contributions

The work was carried out in collaboration between all authors. SYL and HXJ drafted the initial manuscript. XJT and TLP developed the study concept and design. CJ and CLT performed statistical analyses. HJB, MXX, JT, YXH, LF, LJ, BZJ, LJC, WY, FHY, WJ, LYT, BJX, and GXJ provided critical revisions. All authors read and approved the final manuscript.

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.

References

1. Chaudhuri A, Behan P O. Fatigue in neurological disorders[J]. Lancet, 2004, 363(9413): 978-988.
2. Kim S, Jang HJ, Myung W, Kim K, Cha S, Lee H, et al. Heritability estimates of individual psychological distress symptoms from genetic variation[J]. J Affect Disord, 2019, 252: 413-420.
3. Kluger B M, Herlofson K, Chou K L, Lou J S, Goetz C G, Lang A E, et al. Parkinson's disease-related fatigue: A case definition and recommendations for clinical research[J]. Mov Disord, 2016, 31(5): 625-631.
4. Chung K F, Yu Y M, Yeung W F. Correlates of residual fatigue in patients with major depressive disorder: The role of psychotropic medication[J]. J Affect Disord, 2015, 186: 192-197.
5. Skorvanek M, Gdovinova Z, Rosenberger J, Saeedian R G, Nagyova I, Groothoff J W, et al. The associations between fatigue, apathy, and depression in Parkinson's disease[J]. Acta Neurol Scand, 2015, 131(2): 80-87.
6. Lu Y, Qu HQ, Chen FY, Li XT, Cai L, Chen S, et al. Effect of Baduanjin Qigong Exercise on Cancer-Related Fatigue in Patients with Colorectal Cancer Undergoing Chemotherapy: A Randomized Controlled Trial[J]. Oncol Res Treat, 2019, 42(9): 431-439.
7. Grad F P. The Preamble of the Constitution of the World Health Organization[J]. Bull World Health Organ, 2002, 80(12): 981-984.
8. Xue Y, Liu G, Feng Y, Xu M, Jiang L, Lin Y, et al. Mediating effect of health consciousness in the relationship of lifestyle and suboptimal health status: a cross-sectional study involving Chinese urban residents[J]. BMJ Open, 2020, 10(10): e039701.
9. Wang X, Liu J, Wu C, Liu J, Li Q, Chen Y, et al. Artificial intelligence in tongue diagnosis: Using deep convolutional neural network for recognizing unhealthy tongue with tooth-mark[J]. Comput Struct Biotechnol J, 2020, 18: 973-980.
10. Li X, Zhang Y, Cui Q, Yi X, Zhang Y. Tooth-Marked Tongue Recognition Using Multiple Instance Learning and CNN Features[J]. IEEE Trans Cybern, 2019, 49(2): 380-387.
11. 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[J]. Diagnostics (Basel), 2020, 10(7).
12. 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[J]. Endocrine, 2020.
13. Tang AC, Chung JW, Wong TK. Digitalizing traditional chinese medicine pulse diagnosis with artificial neural network[J]. Telemed J E Health, 2012, 18(6): 446-453.
14. Hu MC, Cheng MH, Lan KC. Color Correction Parameter Estimation on the Smartphone and Its Application to Automatic Tongue Diagnosis[J]. J Med Syst, 2016, 40(1): 18.
15. Zhang B, Wang X, You J, Zhang D. Tongue color analysis for medical application[J]. Evid Based Complement Alternat Med, 2013, 2013: 264742.
16. Hu XJ, Zhang L, Xu JT, Liu BC, Wang JY, Hong YL, et al. Pulse Wave Cycle Features Analysis of Different Blood Pressure Grades in the Elderly[J]. Evid Based Complement Alternat Med, 2018, 2018: 1976041.

17. 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[J]. Biomed Res Int, 2018, 2018: 2964816.
18. Wang X, Zhang B, Yang Z, Wang H, Zhang D. Statistical analysis of tongue images for feature extraction and diagnostics[J]. IEEE Trans Image Process, 2013, 22(12): 5336-5347.
19. Kamarudin N D, Ooi C Y, Kawanabe T, Odaguchi H, Kobayashi F. A Fast SVM-Based Tongue's Colour Classification Aided by k-Means Clustering Identifiers and Colour Attributes as Computer-Assisted Tool for Tongue Diagnosis[J]. J Healthc Eng, 2017, 2017: 7460168.
20. Zhang JF, Xu JT, Hu XJ, Chen QG, Tu , Huang JB, et al. Diagnostic Method of Diabetes Based on Support Vector Machine and Tongue Images[J]. Biomed Res Int, 2017, 2017: 7961494.
21. Ding T, Feng L, Rong L, Xi LD. Tongue inspection on Fatigue [C]. The 10th annual Conference of Rehabilitation Committee of Traditional Chinese Medicine of China Disabled Persons' Rehabilitation Association, 2015: 4.
22. Li WL, Yi ZX, Min P. Objective analysis of complexion and tongue color in patients with chronic fatigue syndrome[J]. Shandong Medical Journal, 2019, 59(05): 81-83.
23. Xu JT, Bao YM, Gong BM. Experimental Study on Evaluation of Sphygmogram of Chronic Motion Fatigue [J]. Shanghai Journal of Traditional Chinese Medicine, 2008, (09): 42-44.
24. Kung YY, Kuo T B J, Lai C T, Shen Y C, Su Y C, Yang C C H. Disclosure of suboptimal health status through traditional Chinese medicine-based body constitution and pulse patterns[J]. Complement Ther Med, 2020: 102607.
25. Shi HZ, Fan QC, Gao JY, Liu JL, Bai GE, Mi T, et al. Evaluation of the health status of six volunteers from the Mars 500 project using pulse analysis[J]. Chin J Integr Med, 2017, 23(8): 574-580.
26. Li S. [Network target: a starting point for traditional Chinese medicine network pharmacology][J]. Zhongguo Zhong Yao Za Zhi, 2011, 36(15): 2017-2020.
27. Liu ZH, Sun XB. [Network pharmacology: new opportunity for the modernization of traditional Chinese medicine][J]. Yao Xue Xue Bao, 2012, 47(6): 696-703.
28. Wang ZF, Hu YQ, Wu QG, Zhang R. Virtual Screening of Potential Anti-fatigue Mechanism of Polygonati Rhizoma Based on Network Pharmacology[J]. Comb Chem High Throughput Screen, 2019, 22(9): 612-624.
29. Liu H, Zeng L, Yang K, Zhang G. A Network Pharmacology Approach to Explore the Pharmacological Mechanism of Xiaoyao Powder on Anovulatory Infertility[J]. Evid Based Complement Alternat Med, 2016, 2016: 2960372.
30. Wu L, Gao X, Cheng Y, Wang Y, Zhang B, Fan X. [Symptom-based traditional Chinese medicine slices relationship network and its network pharmacology study][J]. Zhongguo Zhong Yao Za Zhi, 2011, 36(21): 2916-2919.
31. Zhang R, Zhu X, Bai H, Ning K. Network Pharmacology Databases for Traditional Chinese Medicine: Review and Assessment[J]. Front Pharmacol, 2019, 10: 123.

32. Peckham A D, Jones P, Snorrason I, Wessman I, Beard C, Björgvinsson T. Age-related differences in borderline personality disorder symptom networks in a transdiagnostic sample[J]. J Affect Disord, 2020, 274: 508-514.
33. Song J, Liu X, Deng Q, Dai W, Gao Y, Chen L, et al. A network-based approach to investigate the pattern of syndrome in depression[J]. Evid Based Complement Alternat Med, 2015, 2015: 768249.
34. Shi Q, Zhao H, Chen J, Ma X, Yang Y, Zheng C, et al. Study on TCM Syndrome Identification Modes of Coronary Heart Disease Based on Data Mining[J]. Evid Based Complement Alternat Med, 2012, 2012: 697028.
35. Chen J, Lu P, Zuo X, Shi Q, Zhao H, Luo L, et al. Clinical data mining of phenotypic network in angina pectoris of coronary heart disease[J]. Evid Based Complement Alternat Med, 2012, 2012: 546230.
36. Henry T R, Marshall S A, Avis N E, Levine B J, Ip E H. Concordance networks and application to clustering cancer symptomology[J]. PLoS One, 2018, 13(3): e0191981.
37. Fernandez-Rodriguez J, Moser F, Song M, Voigt C A. Engineering RGB color vision into Escherichia coli[J]. Nat Chem Biol, 2017, 13(7): 706-708.
38. Schiller F, Valsecchi M, Gegenfurtner K R. An evaluation of different measures of color saturation[J]. Vision Res, 2018, 151: 117-134.
39. Sun X, Young J, Liu JH, Bachmeier L, Somers R M, Chen K J, et al. Prediction of pork color attributes using computer vision system[J]. Meat Sci, 2016, 113: 62-64.
40. Zhu T, Zhang SP, Guo RX, Chang GC. Improved evaluation method for node importance based on node contraction in weighted complex networks[J]. Systems Engineering and Electronics, 2009, 31(08): 1902-1905.
41. Tan YJ, Wu J, Deng HZ. Evaluation Method for Node Importance based on Node Contraction in Complex Networks[J]. Systems Engineering Theory Practice, 2006, (11): 79-83+102.
42. Enoka R M, Duchateau J. Translating Fatigue to Human Performance[J]. Med Sci Sports Exerc, 2016, 48(11): 2228-2238.
43. Yuan YH. Relationship between tongue image and liver function in virus hepatitis patients – a report of 200 cases[J]. Jiangsu Journal of Traditional Chinese Medicine, 2003, 24(01): 12.
44. Lin RY, Yu HY, Qin JY, Li YY, Wang YH, Yang YZ, et al. Association between tongue coating thickness and clinical characteristics among idiopathic membranous nephropathy patients[J]. J Ethnopharmacol, 2015, 171: 125-130.

Figures

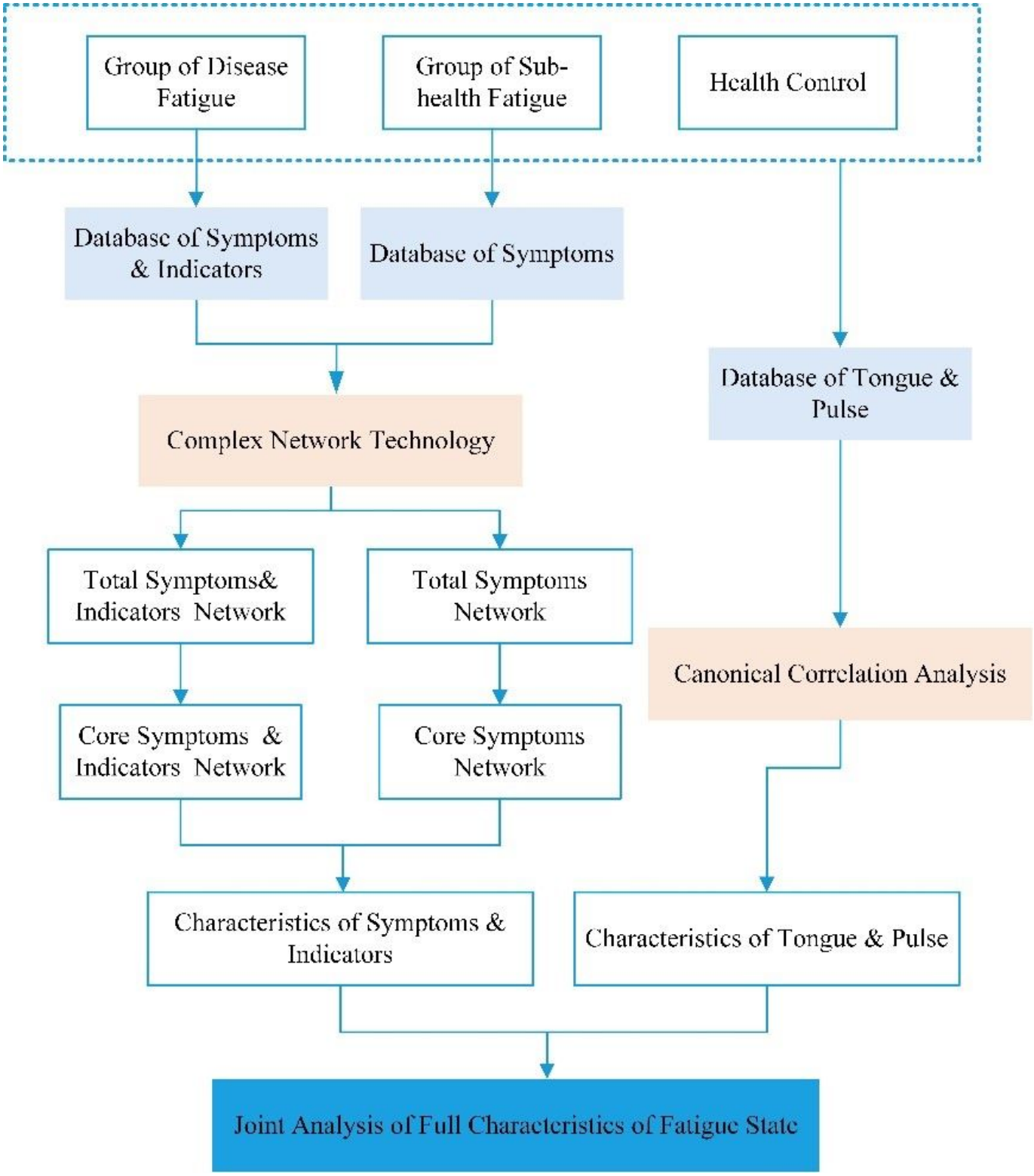

Figure 1

Overall flow diagram.

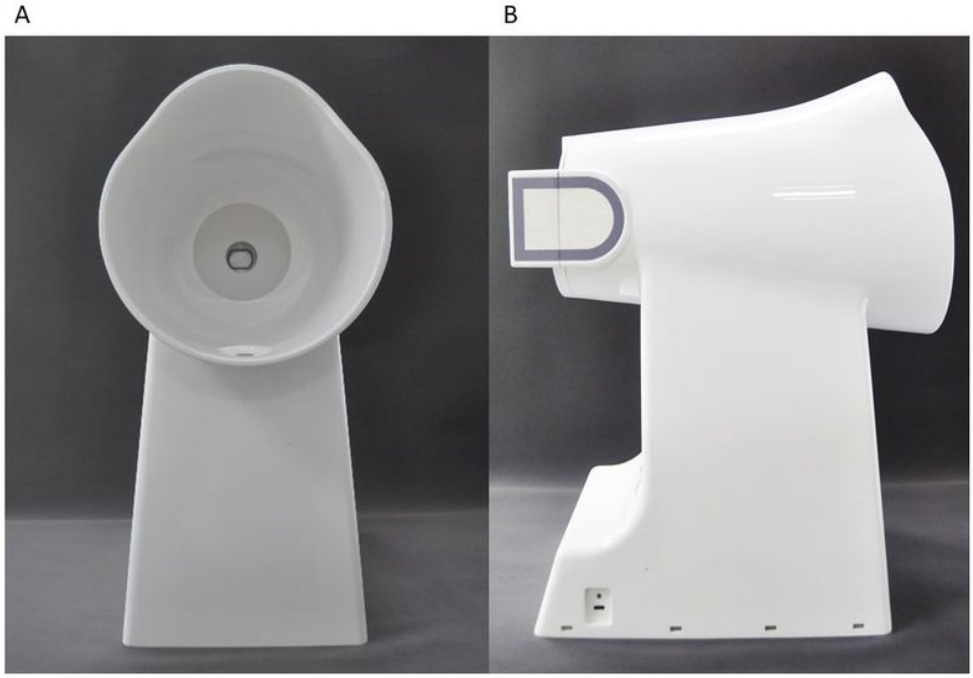

Figure 2

Figures of TFDA-1 tongue diagnosis instrument. (A): Front view.(B): Profile view.

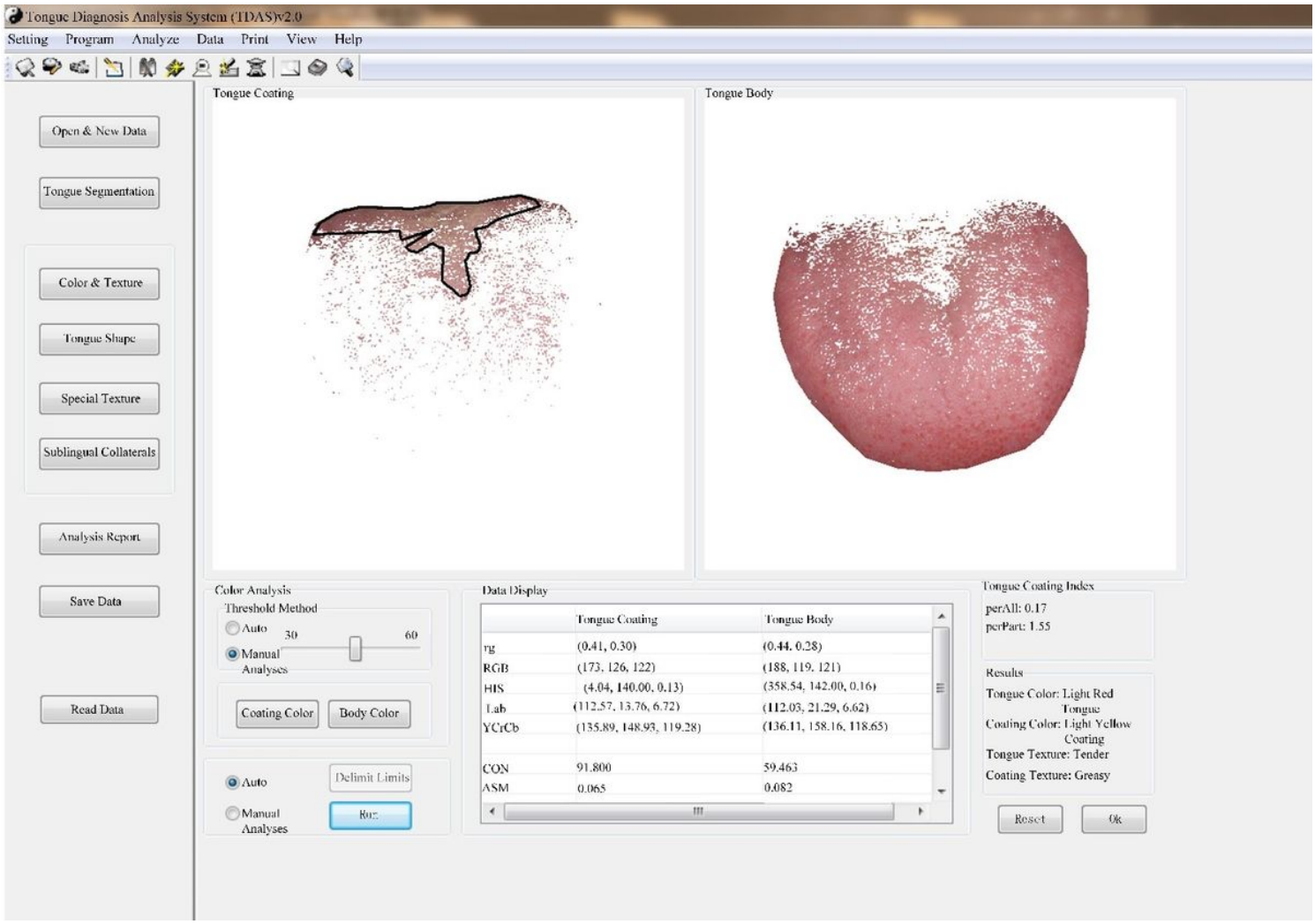

Figure 3

Tongue image analysis interface.

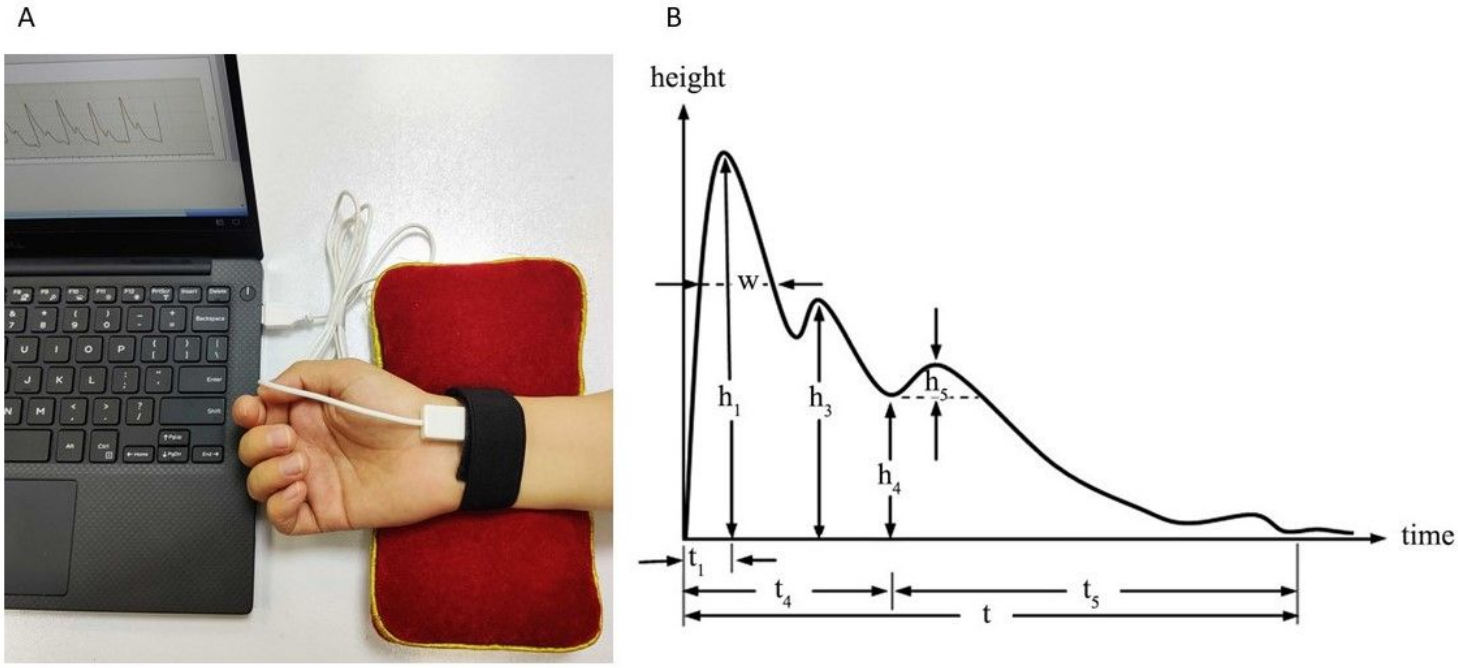

Figure 4

Figures of PDA-1 pulse diagnosis instrument and sphygmogram. (A): PDA-1 pulse diagnosis instrument and supporting equipments.(B): Sphygmogram of PDA-1 pulse diagnosis instrument.

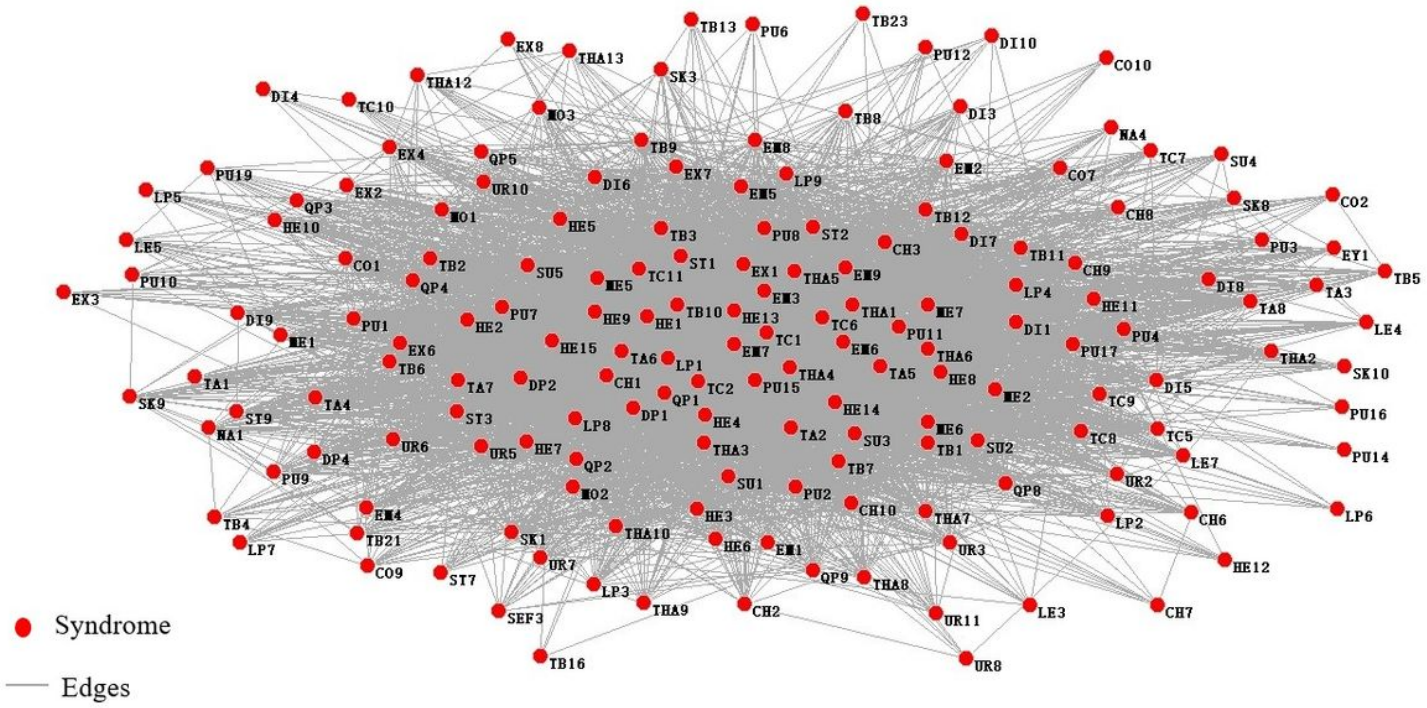

Figure 5

Symptom network of the group of sub-health fatigue.

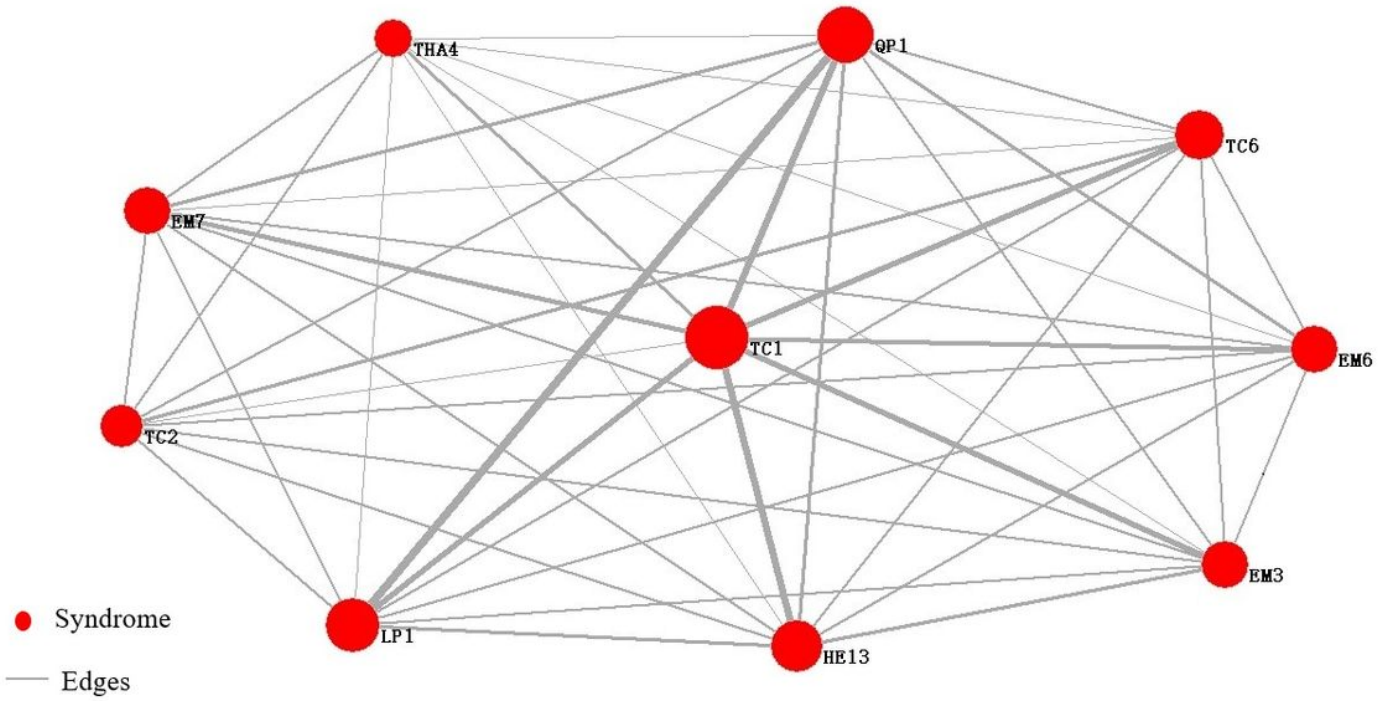

Figure 6

Network of core symptom of the group of sub-health fatigue. TC1:white tongue coating; LP1:headache; TC2:yellow tongue coating; QP1:sour; EM7:dreaminess; EM3:irritability; THA4:chest distress; HE13:xerophthalmia; TC6:thick coating; EM6:insomnia.

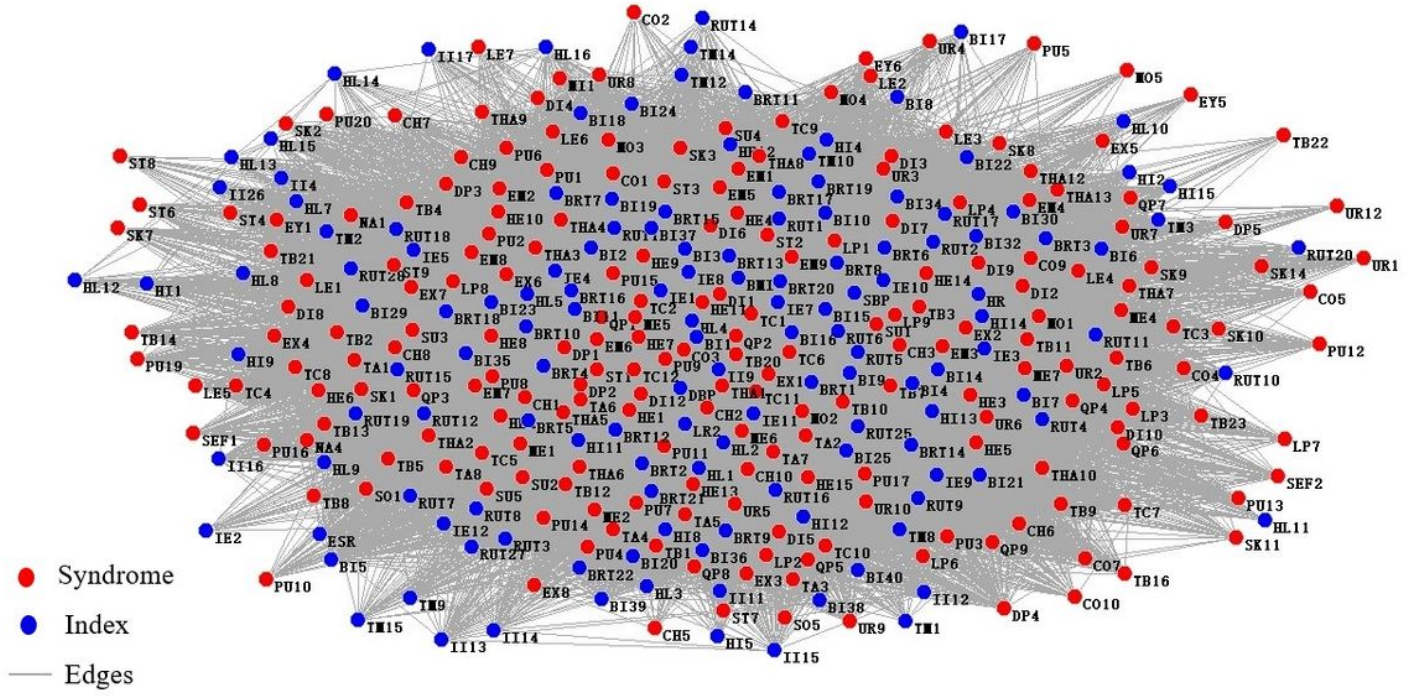

Figure 7

Symptom and index network of the group of disease fatigue.

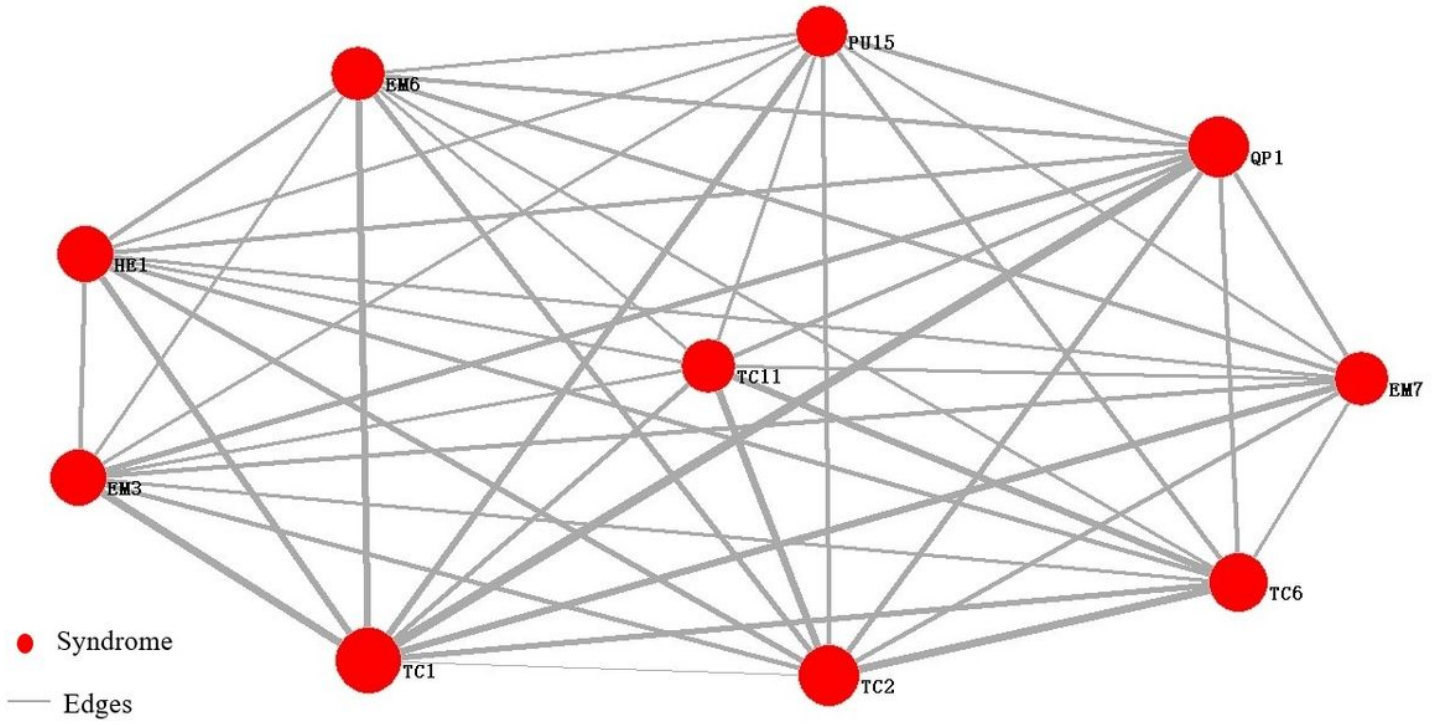

Figure 8

Network of core symptom of the group of disease fatigue. TC1:white tongue coating; HE1:dizziness; TC2:yellow tongue coating; QoP1:sour; EM7:dreaminess; TC6:thick coating; PU15:string-like pulse; TC11:greasy coating; EM6:insomnia; EM3:irritability.

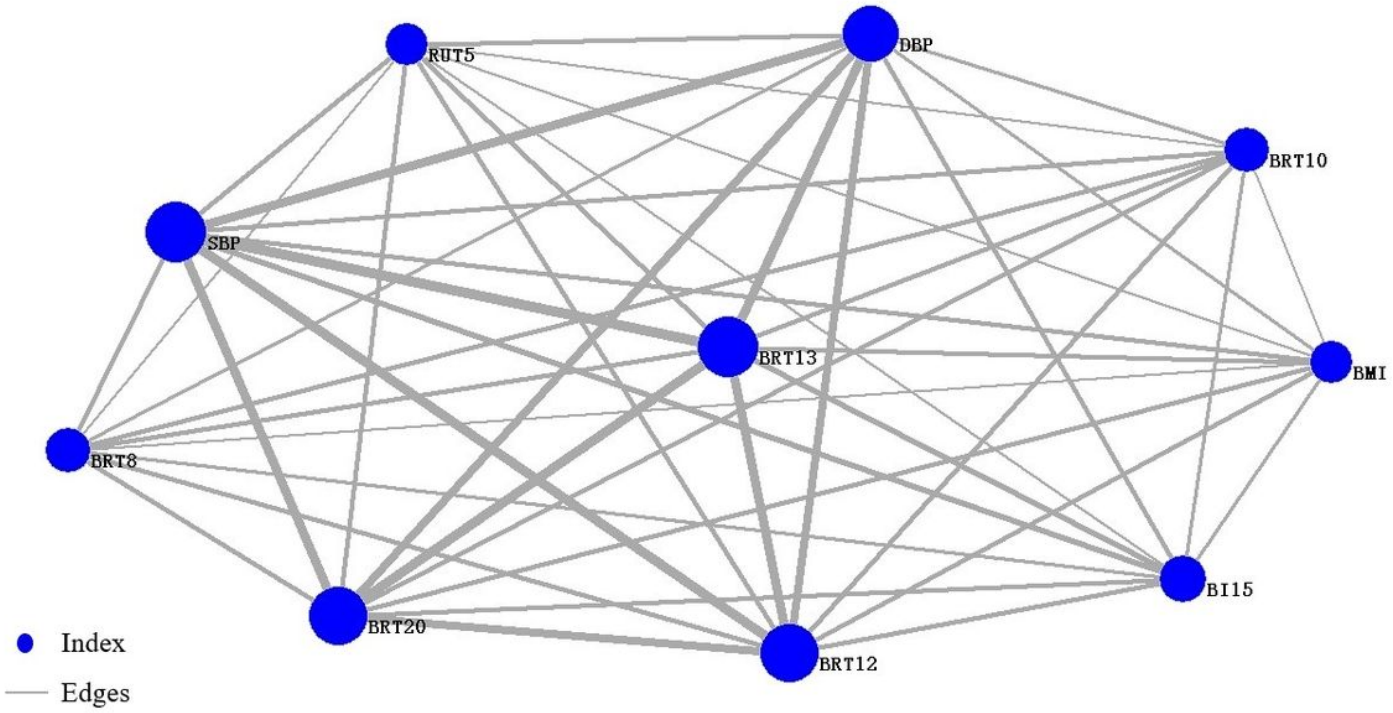

Figure 9

Network of core index of the group of disease fatigue. BRT13:basophil; BRT20:platelet distribution width; SBP:systolic blood pressure; BRT12:percentage of monocyte; DBP :diastolic blood pressure; RUT5:PH of urine; BRT8:hemoglobin; BRT10:hematocrit; BI15:uric acid; BMI :Body mass index.

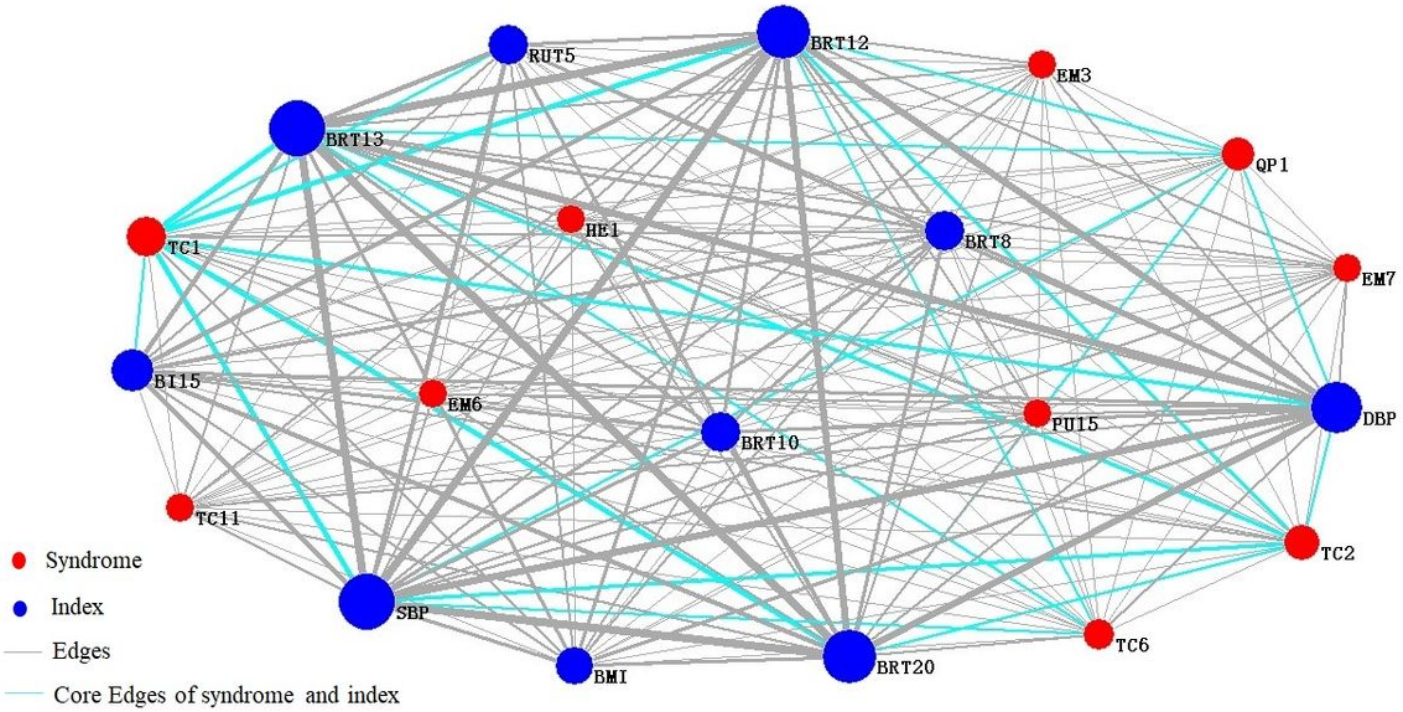

Figure 10

Network of core symptom-indicator of the group of disease fatigue. TC1:white tongue coating; HE1:dizziness; TC2:yellow tongue coating; QoP1:sour; EM7:dreaminess; TC6:thick coating; PU15:string-like pulse; TC11:greasy coating; EM6:insomnia; EM3:irritability; BRT13:basophil; BRT20:platelet distribution width; SBP:systolic blood pressure; BRT12:percentage of monocyte; DBP :diastolic blood pressure; RUT5:PH of urine; BRT8:hemoglobin; BRT10:hematocrit; BI15:uric acid; BMI:body mass index.

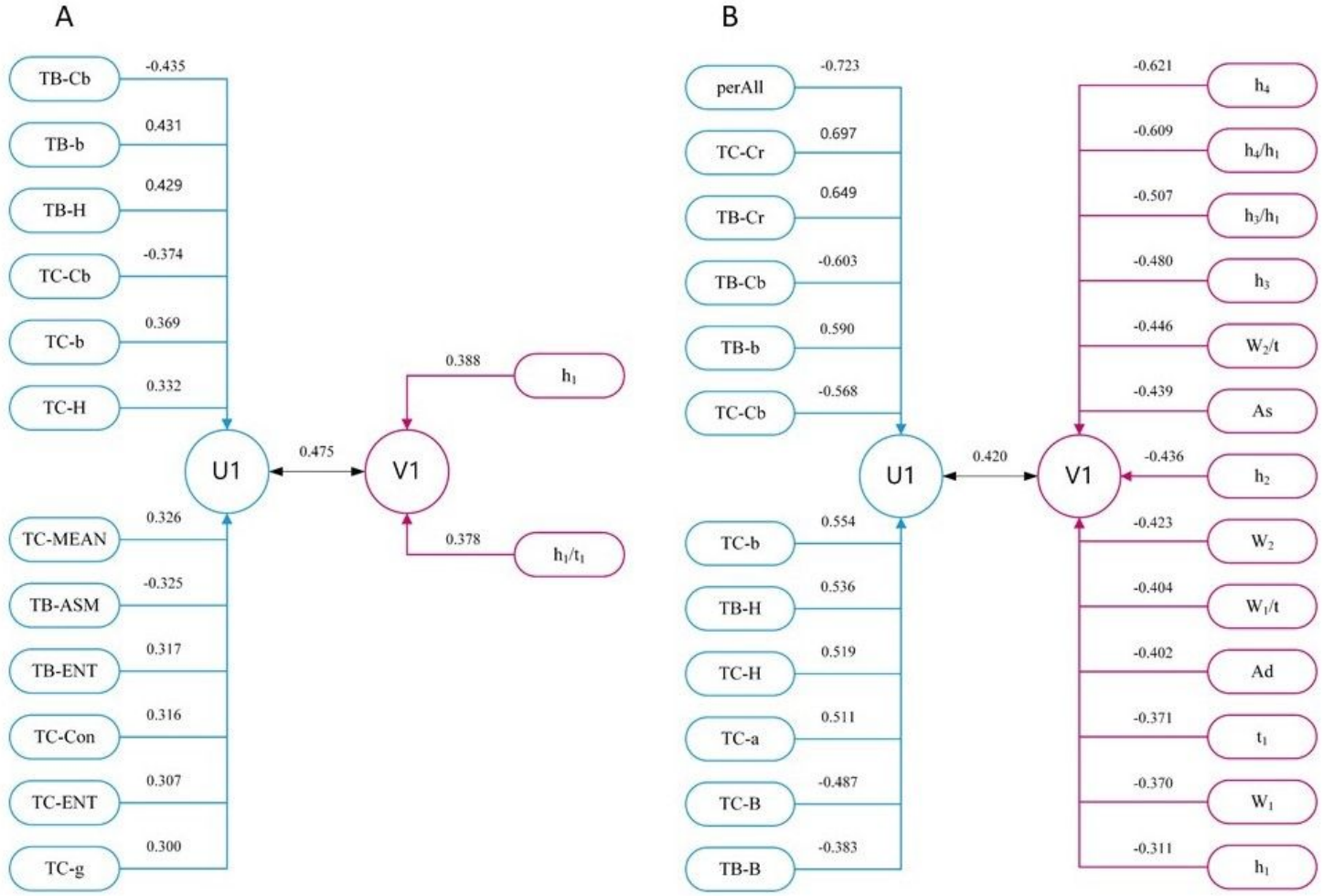

Figure 11

Structure diagram of canonical correlation analysis of tongue and pulse parameters. The left indexes of Figure A and Figure B are the parameters of tongue, and the right indexes are the parameters of pulse. The prefix TB represents the tongue body index, and prefix TC represents the tongue coating index.U1 is the representative comprehensive variable extracted from the tongue parameters, V1 is the representative comprehensive variable extracted from the pulse parameters.(A): Healthy control group.(B): The group of disease fatigue.