Int J Clin Exp Med 2016;9(11):21888-21899
www.ijcem.com /ISSN:1940-5901/IJCEM0035172

Original Article

Classification and differentiation between kidney yang and yin decficiency syndromes in TCM based on decision tree analysis method

Tieniu Zhao[1], Huijun Wang[1], Chunquan Yu[1], Jing Wang[2], Yuanwu Cui[3], Xia Zheng[4], Bin Wang[5], Wenjuan Wang[6], Jingyan Meng[1]

[1]College of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 300193, PR China; [2]Spine Research Institute, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai 200032, PR China; [3]Center of Moxibustion, Second Affiliated Hospital of Tianjin University of Traditional Chinese Medicine, Tianjin 300150, PR China; [4]Department of Traditional Chinese Medicine, Second Affiliated Hospital of Chengdu University of Traditional Chinese Medicine, Chengdu 610042, PR China; [5]Department of Male, Dongzhimen Hospital Affiliated to Beijing University of Traditional Chinese Medicine, Beijing 100007, PR China; [6]Department of Traditional Chinese Medicine, Capital Medical University, Beijing 100069, PR China

Received July 6, 2016; Accepted September 20, 2016; Epub November 15, 2016; Published November 30, 2016

Abstract: The classification and differentiation between kidney Yang and Yin deficiency syndromes (KDS-YANG and KDS-YIN) in traditional Chinese medicine (TCM) was performed with the approaches of decision tree as well as logistic regression analysis. A clinical epidemiology study on 44 symptoms, 5 tongue signs and 2 pulse signs (TCM diagnostic features) was conducted among 2,765 patients with KDS. Accordingly, an effective syndrome-identifying model on KDS-YANG and KDS-YIN was established. The results indicated that in terms of statistical significance ($P < 0.05$) of the cases with the above two different syndromes, the diagnostic indicators such as 11 symptoms, 2 tongue signs, and 1 pulse signs was found out by logistic regression analysis. The accurate rate of differentiating KDS-YANG and KDS-YIN was proved to be 88.0%. What's more, four symptoms-aversion to cold, cold limbs, pale tongue, deep and thread pulse-were confirmed to be the most significant variables, which was beneficial for the syndrome differentiation. The decision tree analysis model was an effective approach to differentiate KDS-YANG and KDS YIN, which could be helpful to change the syndrome-diagnosing method from experience-based to data-model based.

Keywords: Syndrome differentiation, kidney deficiency syndrome, kidney yin deficiency syndrome, kidney yang deficiency syndrome, decision tree

Introduction

Traditional Chinese medicine (TCM) is a scientific study on disease etiology and pathogenesis based on Yin and Yang theory as well as the five elements theory; therefore syndromes are regarded as the study objects. According to TCM theory, syndrome are known as "patterns" or "zhengs (Chinese pinyin)" and serve as the key concept and theoretical abstract of the symptom profiles of the disease diagnosis and treatment in China for over 2,500 years [1, 2]. Syndromes or TCM zhengs are just like a vivid picture created by the clinical manifestations that could clearly show the features, the locations and the pathology conditions of the diseases, which have been found out through comprehensive analysis of all symptoms and signs based on four fundamental TCM diagnostic methods, that is, observation, auscultation and olfaction, inquiry as well as pulse-taking and palpation [3-5]. In TCM, kidney deficiency syndrome (KDS) is mostly of asthenia nature and usually attributes to the constitutional deficiency, the vital essence insufficiency during childhood, the essence exhaustion for the seniors, the intemperance of sexual life, and the involvement of kidney in other viscera disorders, which leads to the deficiency of yin, yang, essence and Qi. Kidney yin deficiency syndrome (KDS-YIN) is manifested as the symptoms of endogenous

asthenic heat due to the consumption of kidney yin and the nourishment insufficiency [6]. KDS-YIN, which is a general term describing the deficient condition of kidney system, is mail manifested as: lower back pain, soreness and weakness of the lumbar region and knees, dizziness, fatigue, night sweating, hot sensation in the palms, soles and chest, dysphoria, insomnia, thread pulse, reddish tongue with less fur or with white fur [7]. Kidney yang deficiency syndrome (KDS-YANG) refers to the asthenia cold symptoms due to failure of Qi transformation caused by the decline of kidney yang and its failure to nourish the body [6]. It is primarily manifested as the symptoms of soreness and weakness of the lumbar regions and knees, aversion to cold, cold limbs, less libido; the secondary symptoms of low spirit and weakness, hypodynamia, impotence, infertility, frequent and increasing urination at night, clear urine, hair loss, loose teeth, a pale tongue and a deep and thready pulse.

The TCM syndromes may vary among the patients suffering from the same disease because of their individual difference. Thus they are managed by TCM syndrome-specific therapies [8]. In fact, TCM practitioners diagnosed and treated diseases based on the TCM theory of syndrome differentiation. So this is the reason why syndrome differentiation played an important role in the disease identification and therapeutic procession, even ultimately influenced the clinical efficacy of the disease treatment [9]. In other words, a symptom is an observable disorder indicator while a syndrome is a relatively abstract perception which can only be reflected by symptoms in TCM. However, the co-relations between symptom and syndrome in TCM are not yet clearly defined, therefore it is of great difficulty to carry out the clinical diagnosis and scientific study on TCM. In addition, TCM practitioners are greatly influenced by their subjective judgment in diagnosing TCM syndromes and therefore their identification consistency was on low level [10, 11]. Moreover, because of the lack of validated and standardized instruments to assess TCM syndrome, objective assessment should be developed in TCM study [12]. To a certain extent, the above-mentioned factors affect the standardized and objective identification of TCM syndrome. Therefore, syndrome differentiation is of great importance for TCM diagnosis.

It is not only necessary but also possible to construct syndrome model based on the diagnosis model data by combination of a variety of analysis methods. So far, a decision tree model was built and was supposed to forecast TCM syndrome accurately, which had been successfully and widely applied to a variety of clinical diseases, including liver cirrhosis [13], coronary heart disease [14] and chronic hepatitis B [15].

In current study, according to its features the variables information in KDS will be described to analyze the complex and mutual relationship between symptom and syndrome. So it is of great interest to compose KDS questionnaires as measurement tools, based on 5 kidney-related diseases: osteoporosis, male infertility, female inferitility, thalassemia and Alzheimer's disease. During this process, expert's opinion and experience as well as Delphi method based on retrieved literature in TCM study are adopted [16]. The study purpose is to identify TCM syndromes: KDS-YIN and KDS-YANG based on 2,765 cases of patients with symptoms and to predict the syndrome's type of the future cases accurately by using logistic regression analysis combined with quick unbiased efficient statistical tree (QUEST) algorithm analysis. Therefore, the data about the patients with KDS-YANG and KDS-YIN would be analyzed during which diagnostic model were established based on a large-sample clinical survey in the study.

Material and methods

Study objects

The cases of 2,765 patients with KDS from 5 TCM hospitals have been collected by the team members to set up the sample database. The specific cases origins were listed as followings: 810 inpatients and outpatients with primary osteoporosis from Longhua Hospital Affiliated to Shanghai University of TCM; 763 patients with male infertility from Dongzhimen Hospital Affiliated to Beijing University of TCM; 160 patients with thalassemia from Guang'anmen Hospital Affiliated to China Academy of Chinese Medical Science; 759 patients with female infertility from Second Hospital affiliated to Chengdu University of TCM as well as 273 patients with Alzheimer's disease from Second Hospital affiliated to Tianjin University of TCM respectively, within the study time from September 2011 to September 2013. The data

set did not only include one certain disease, but also involved the diseases with KDS.

Data collection items included 51 attributes of TCM (that were 44 symptoms, 5 tongue signs and 2 pulse signs), which inquired TCM symptom variables by using all four fundamental TCM diagnostic methods. The symptom variables used in the survey were elements that a TCM doctor would think over when treating patients with KDS. Each data collection team was led by a team leader who was with the title at least as attending physician. Team members have been trained beforehand to ensure the consistency in their judgment of the presence or absence of symptoms.

The average age of 2,765 patients with KDS was (M_e, Q_d): (36, 35) years old, including 1,054 male patients (38.1%) and 1,711 female patients (61.9%). The average age of male patients was (M_e, Q_d): (32, 13) years old, and the average age of female patients was (M_e, Q_d): (52, 37) years old. The syndrome differentiation was further confirmed by two TCM experts, and the case samples were partitioned into 2 syndrome types as follows: 1,465 cases with KDS-YIN, and 1,300 cases with KDS-YANG.

Diagnosis, inclusion, and exclusion criteria

Diagnosis criteria: Diagnostic criteria of osteoporosis: According to the diagnostic criteria for osteoporosis recommended by World Health Organization (WHO): BMD T-score > -1.0 SD: normal bone mass, -2.5 SD ≤ T-score ≤ -1.0 SD: osteopenia, T-score < -2.5 SD: osteoporosis [17]. The patients identified as kidney deficiyy were included based on TCM syndrome differentiation [18, 19].

Diagnostic criteria of infertility: According to the diagnostic criteria for infertility recommended by WHO: the definition of infertility was the failure of the sexually active but non-contraceptive couple to have spontaneous pregnancy within one years (WHO, 1995) [20, 21].

Diagnostic criteria of thalassemia: It referred to Chinese diagnostic criteria: diagnosis and efficacy standard of blood disease [22], which was combined with diagnostic criteria of the type β-thalassemia intermedia and α-thalassemia Hb-H disease. According to neonates Hb-art content the disease diagnosis would be classified: protein accounted for 1-3% was diagnosed as α_1 thalassemia, that for 3-15% was as α_1 thalassemia, that for 15-40% was as thalassemia intermedia (Hb H disease) and that > 40% was as Hb-Bart hydrops syndrome.

Diagnostic criteria of Alzheimer's disease: It was in reference to the criteria provided by the National Institute of Neurological and Communicative Disorders and Stroke-Alzheimer's Disease and Related Disorders Association (NINCDS-ADRDA) [23], the 4th edition of the Diagnostic and Statistical Manual of Mental Disorders (DSM-IV) [24], and the 10th revision of the International Classification of Diseases (ICD-10) [25]. Of these criteria sets, the NINCDS-ADRDA criteria were most widely used in dementia study.

Inclusion criteria: Patients who met the diagnostic criteria for osteoporosis, male infertility, thalassemia, female infertility and Alzheimer's disease were included. The patients were voluntary to participate in the clinical study and would sign informed consent form.

The exclusion criteria were as follows: (1) Patients who failed to meet western medicine and TCM diagnostic standards; (2) Those who were presented with concomitant diseases, such as cerebrovascular, gastrointestinal, kidney, lung, endocrine, and other serious diseases or mental illness; (3) Those who had abnormal liver and kidney function tests were shown as abnormal; (4) The women who were in pregnancy or breast-feeding; (5) Patients who had severe dysfunction in heart, liver or kidney.

The differentiation criteria of KDS were based on "*The Guiding Principles of Clinical Research of New Chinese Medicine (trial version*)" (published by China pharmaceutical technology publishing house, 2002) [18] and "*Reference Standard of TCM Deficiency Syndrome*" revised by the Chinese integrated traditional and western medicine deficiency syndrome and professional committee of the elderly in 1986 to develop a "Reference standard TCM deficiency syndrome" [19]. Two basic syndromes of TCM were set up as follows.

The primary symptoms of KDS-YIN were manifested as soreness and weakness of the lumbar regions of knees, hot sensation of palms,

Table 1. Clinical questionnaire of symptom and its value of KDS in TCM

ID	Symptom	Evaluation	ID	Symptom	Evaluation
1	Dull expression	None or yes	27	Cold limbs	None or yes
2	Spiritlessness	None or yes	28	Shortness of breath	None or yes
3	Hypodynamia	None or yes	29	Soreness and weakness of the lumbar region and knees	None or yes
4	Withered hair/White hair/Hair loss	None or yes	30	Soft bone	None or yes
5	Loose teeth/tooth removal	None or yes	31	Tidal fever	None or yes
6	Pale complexion	None or yes	32	Dysphoria hot/Feverish palms and soles	None or yes
7	Pale white complexion	None or yes	33	Dry stool	None or yes
8	Bright white complexion	None or yes	34	Loose stool	None or yes
9	Dim complexion	None or yes	35	Yellow urine and oliguria	None or yes
10	Pale lips	None or yes	36	Clear urine	None or yes
11	Reddish face	None or yes	37	Frequent micturition	None or yes
12	Reheat	None or yes	38	Frequent and increased urination at night	None or yes
13	Amnesia	None or yes	39	Growth retardation	None or yes
14	Slow reaction	None or yes	40	Less libido	None or yes
15	Dry mouth	None or yes	41	Impotence	None or yes
16	Dry pharynx	None or yes	42	Spermatorrhea	None or yes
17	Poor appetite and less amount of food	None or yes	43	Male infertility	None or yes
18	Dizziness	None or yes	44	Female infertility	None or yes
19	Tinnitus	None or yes	45	Pale tongue	None or yes
20	Deafness	None or yes	46	Reddish tongue	None or yes
21	Cardiopalmus/Palpitation	None or yes	47	White fur	None or yes
22	Dysphoria	None or yes	48	Thin and white fur	None or yes
23	Insomnia	None or yes	49	Less fur	None or yes
24	Dreaminess	None or yes	50	Deep and thready pulse	None or yes
25	Night sweating	None or yes	51	Thready and rapid pulse	None or yes
26	Aversion to cold	None or yes			

soles and chest. The secondary symptoms as dizziness, tinnitus or deafness, insomnia, dreaminess, tidal fever, night sweating, dry stools, dry mouth and throat, a reddish tongue, a thread rapid pulse.

KDS-YANG is mainly manifested as the primary symptoms of soreness and weakness of the lumbar regions and knees, aversion to cold, cold limbs, less libido; secondary symptoms of spiritlessness and weakness, hypodynamia, impotence, infertility, frequent and increased urination at night, clear urine, hair loss, loose teeth, a pale tongue and a deep and thready pulse.

Data management and statistical analysis

Data management was performed by using software of Epidata 3.1. The data was arranged in a column wise format with each subject given a sequence identifier. Data with normal distribution was expressed as mean ± SD for continuous variable, while data with no normal distribution was expressed as (M_e, Q_d) for continuous variable. Data was expressed as numbers with corresponding percentages for categorical variables. Differences in proportions were assessed by using a χ^2-test. Comparisons of the two groups were performed via Mann-Whitney U test. The method of logistic regression analysis in combination with quick unbiased efficient statistical tree (QUEST) algorithm analysis was used in the study. 51 variables with multivariate were analyzed to compare the differences between KDS-YIN and KDS-YANG. These 51 variables were defined as independent variables while KDS as a dependent variable. They were examined in a multivariate model by using forward stepwise maximum likelihood logistic regression to identify the symptoms of KDS (α=0.05). Odds ratios (ORs) were estimated by multivariate logistic regression analysis. As shown in **Table 1**, the 51 variables were finally collected to construct the complete clinical questionnaires of symptom set (feature set as below) of KDS in TCM. All reported *P* values were those of two-sided tests. The statistical significance was set at $P < 0.01$. The statistical algorithm that selected variables and quick unbiased efficient statistical tree (QUEST) algorithm analysis was used

Table 2. Comparison of sociodemographic characteristics between KDS-YANG and KDS-YIN

Total	KDS				χ^2 value	*P* value
	KDS-YANG (n=1300)		KDS-YIN (n=1465)			
Gender	n	(%)	n	(%)	8.223	0.004
Male	459	35.3	595	40.6		
Female	841	64.7	870	59.4		
Age (years)					4.942	0.085
< 40	769	59.2	821	56.0		
40-60	143	11.0	199	13.6		
> 60	388	29.8	445	30.4		
(M_e, Q_d)	(35, 33)		(35, 34)*			0.224
Education (years)					18.481	< 0.001
-6	157	12.1	248	16.9		
-12	544	41.8	638	43.5		
≥ 13	599	46.1	579	39.5		
Occupation					1.270	0.260
White collar	809	62.2	942	64.3		
Blue collar	491	37.8	523	35.7		
Ethnic					44.980	< 0.001
HAN	1269	97.6	1345	91.8		
Other	31	2.4	120	8.2		
Marital status					48.550	< 0.001
Married	1258	96.8	1320	90.1		
Unmarried	42	3.2	145	9.9		

M_e: Median, Q_d: Quartile deviation, *U=923514.5, *P*=0.224.

to develop the decision tree models. QUEST decision tree was nonparametric procedure that made no assumptions of the underlying data. This algorithm determined how categorical independent variables best can be combined without bias to predict a binary outcome based on "if-then" logic and to build accurate binary trees quickly and efficiently. Data was analyzed by using the statistical software of SPSS version 20.0.

Results

Sociodemographic characteristics between KDS-YANG and KDS-YIN

In the 2,765 cases under investigation, 1,465 cases were diagnosed as KDS-YIN while 1,300 cases as KDS-YANG by means of KDS differentiation criteria used by experts. As shown in **Table 2**, among the 1,465 cases of KDS-YIN, 595 patients (40.6%) were male and 870 (59.4%) were female, age (M_e, Q_d): (35, 34). Among 1,300 cases of KDS-YANG, 459 patients (35.3%) were male and 841 (64.7%) were female, age (M_e, Q_d): (35, 33). Compared with KDS-YANG, lower proportions of KDS-YIN were females (59.4% versus 64.7%, P < 0.01), the proportions of KDS-YIN were younger patients (56.0% versus 59.2%, P > 0.05), lower proportions of KDS-YIN were with higher education (39.5% versus 46.1%, P < 0.01), the proportions of KDS-YIN were white-collar employees (64.3% versus 62.2%, P=0.260), higher proportions of KDS-YIN were Han ethnics (91.8% versus 97.6%, P < 0.01), higher proportions of KDS-YIN were the married (90.1% versus 96.8%, P < 0.01).

Multivariate logistic regression analysis of relevant symptoms for KDS

As shown in **Table 3**, logistic regression analysis showed that reddish face, dry mouth, night sweating, tidal fever, feverish palms and soles, dry stools and a thready and rapid pulse were relevant symptoms for KDS-YIN. On the contrary, the symptoms, such as aversion to cold, cold limbs, loose stools, clear urine, impotence, pale tongue and white fur were related with KDS-YANG. The most significant symptoms of the differences between KDS-YANG and KDS-YIN were reddish face, dry mouth, night sweating, aversion to cold, cold limbs, feverish palms and soles, loose stools, clear urine, a pale tongue with white fur and a thready and rapid pulse (P < 0.01). As shown in **Table 4**, the results based on 2,765 cases of patients with symptoms. They were generated by the logistic regression model, which showed 2,578 cases of patients were accurately classified. The diagnostic accuracy was 94.5%. The sensitivity was 94.7%. The specificity was 94.2%.

QUEST algorithm of decision tree analysis: establishment of identification model on KDS and validation

At first, data standardization was used to analyze the information of KDS in the cases from different hospitals. Next, the identification models of KDS were constructed by QUEST decision tree, one of algorithm of decision tree

Table 3. Symptoms associated with use of traditional chinese medicine in KDS patients in multivariate logistic regression model

Variable	B	S.E.	Wald	df	*P*	Exp (B)	95% CI
Reddish face	1.136	0.357	10.143	1	0.001	3.115	(1.548, 6.269)
Dry mouth	1.058	0.195	29.590	1	0.001	2.881	(1.968, 4.219)
Night sweating	1.342	0.254	27.978	1	0.001	3.826	(2.327, 6.291)
Aversion to cold	-2.654	0.211	158.930	1	0.001	0.070	(0.047, 0.106)
Cold limbs	-1.886	0.212	78.822	1	0.001	0.152	(0.100, 0.230)
Tidal fever	0.879	0.316	7.753	1	0.005	2.409	(1.297, 4.472)
Feverish palms and soles	1.961	0.256	58.486	1	0.001	7.103	(4.298, 11.740)
Dry stool	0.663	0.230	8.322	1	0.004	1.941	(1.237, 3.046)
Loose stool	-0.959	0.287	11.153	1	0.001	0.383	(0.218, 0.673)
Clear urine	-1.866	0.370	25.453	1	0.001	0.155	(0.075, 0.319)
Impotence	-1.239	0.423	8.558	1	0.003	0.290	(0.126, 0.664)
Pale tongue	-1.581	0.199	63.391	1	0.001	0.206	(0.139, 0.304)
White fur	-0.826	0.230	12.900	1	0.001	0.438	(0.279, 0.687)
Thready and rapid pulse	1.135	0.212	28.807	1	0.001	3.112	(2.056, 4.711)

Table 4. Multivariate logistic regression results of classification for 2,765 cases

Multivariate logistic regression	TN FN	FP TP	Sensitivity (%)	Specificity (%)	Accuracy (%)
KDS	1206	74	94.7	94.2	94.5
	77	1372			

Note: Sensitivity = TP/(TP+FN); Specificity = TN/(TN+FP); Accuracy = (TP+TN)/(TP+FN+TN+FP).

analysis. KDS was considered as dependent variable whereas 14 attributes of TCM (including 11 symptoms, 2 tongue signs and 1 pulse signs) were labeled as independent variables. "Parent Node" 100 and "Child Node" 50 were set up which allowed the tree model to develop sufficiently. Split-sample validation was applied in the study to minimize the bias, which was produced by random sampling of the training sample (75%) and test sample (25%). The decision tree algorithm classified the data into subgroups with statistical significance that was exclusive and exhaustive for both sides [26]. To increase the operability of clinical use, the number of branches of the decision tree would be limited to 4. As shown in **Figure 1**, in this model, the tree analysis showed the 4-level QUEST decision tree with a total of 9 nodes, of which 5 were terminal nodes.

Four major predictors of symptoms which reached significance and were included in this model were demonstrated as aversion to cold, cold limbs, a pale tongue, and a deep and thready pulse. The other 42 symptoms, 4 tongue signs and 1 pulse sign were not significant in the model. As shown in **Table 5**, the diagnostic model used to differentiate these two types of KDS among 2,138 cases of the training sample had an overall accurate diagnostic rate of 89.9%, with the sensitivity of 83.8% and specificity of 96.8%. These two types of KDS of test sample for 627 cases had an overall accurate diagnostic rate of 88.0%, with the sensitivity at 80.7% and specificity at 96.0% respectively.

The first level of the QUEST decision tree was split into two initial branches in terms of the first level on cold limbs. The symptom of cold limbs was the best symptom to identify KDS-YANG, therefore the classification accuracy of KDS-YANG was at 91.7%. Otherwise, 82.3% patients with no cold limbs were identified as KDS-YIN.

As seen in the second level of the QUEST decision tree, aversion to cold was shown to be the next best predictor variable for cases without cold limbs. The classification accuracy of KDS-YANG was 60.6% for patients with aversion to cold. Otherwise, accurate diagnostic rate of KDS-YIN for patients without aversion to cold was 90.2%.

A pale tongue was the most prominent variable on the third level of the QUEST decision tree. For those cases without pale tongue, the accu-

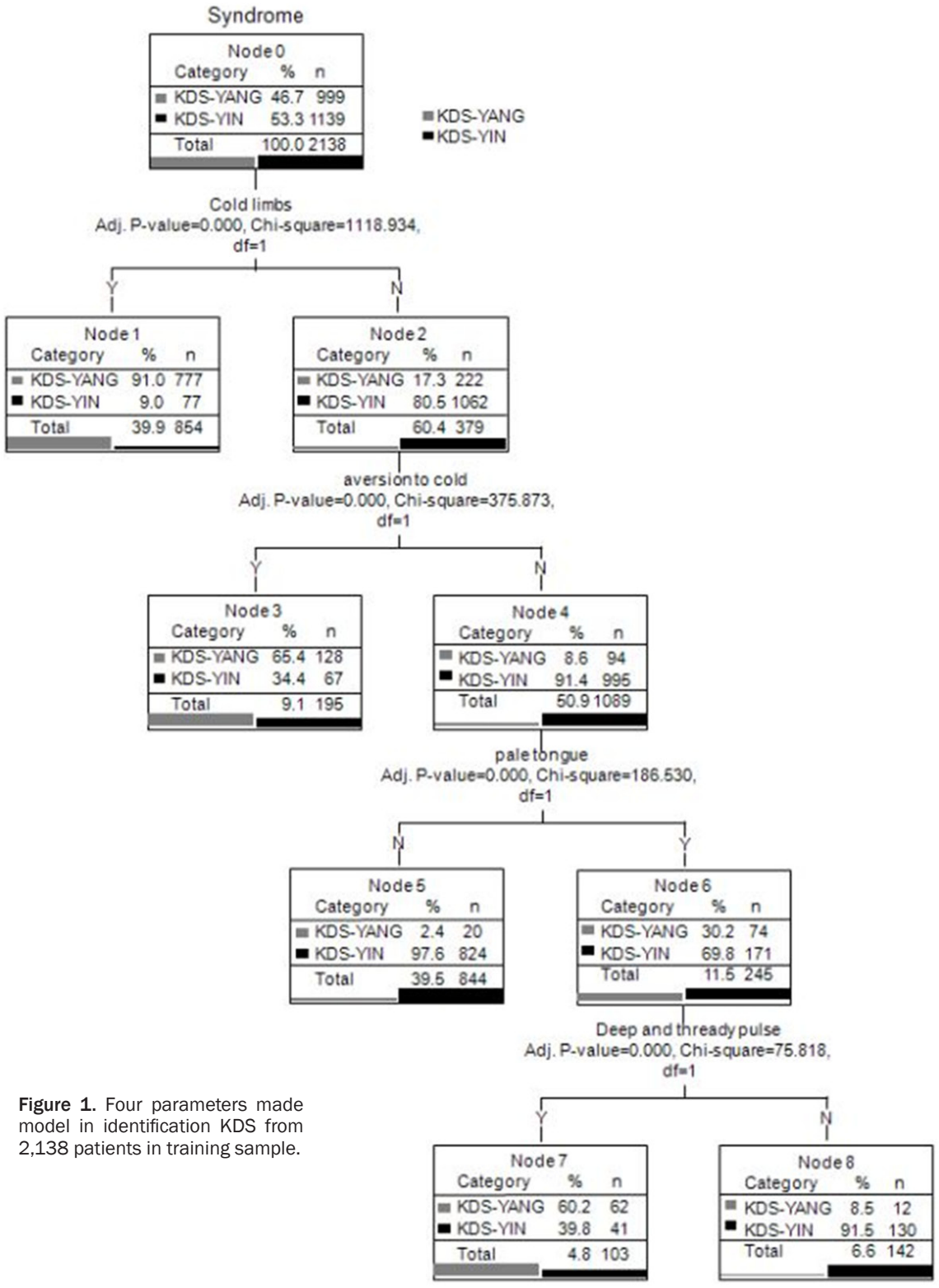

Figure 1. Four parameters made model in identification KDS from 2,138 patients in training sample.

rate identification rate of KDS-YIN was 97.0%. While the accurate identification rate of KDS-YANG for those with pale tongue was 69.3%.

A deep and thready pulse was the most prominent variable on the fourth level of the QUEST decision tree. The classification accuracy of

Table 5. Split-sample validation results of classification of KDS for 2765 cases

QUEST	N	TN FN	FP TP	Sensitivity (%)	Specificity (%)	Accuracy (%)
Training sample	2138	967	32	83.8	96.8	89.9
		185	954			
Test sample	627	289	12	80.7	96.0	88.0
		63	263			
Sum	2765	1256	44	83.1	96.6	89.4
		248	1217			

Note: Sensitivity = TP/(TP+FN); Specificity = TN/(TN+FP); Accuracy = (TP+TN)/(TP+FN+TN+FP).

KDS-YANG was 56.1% for patients with deep and thready pulse. In contrast, accurate diagnostic rate of KDS-YIN for patients without deep and thready pulse was at 91.5%.

As shown in **Figures 1** and **2**, the results showed that cold limbs were the best predictive variable quantity of KDS among the four variables. However, we could not differentiate between KDS-YANG and KDS-YIN only by cold limbs. We could fall back on either the second grade variable quantity: aversion to cold, or the third grade variable quantity a pale tongue, or the final-grade variable quantity a deep and thready pulse. Due to the complexity of KDS, its study was hard to find the golden index for syndrome differentiation. However, the combinations of different syndromes and signs might demonstrate the characteristics of different KDS. Syndrome diagnosis model was gradually established on the basis of digital information of the four main fundamental TCM diagnostic methods: observation, auscultation and olfaction, inquiry, pulse-taking and palpation, and further verified by clinical practice. It was important to promote objective study on identification of TCM by improving unremittingly research on the diagnosis model of syndromes of TCM.

Discussion

Identification of patterns (also known as syndrome differentiation), which is the essence of Chinese medical diagnosis, refers to the process of identifying the basic disharmony that underlies all clinical manifestations. Since there's no simple corresponding relationship between a symptom and a pathological mechanism, and some symptoms that helps to distinguish a syndrome are not always taken as indispensable, which, making the course of identification of patterns more difficult and error-prone. So the syndrome differentiation between kidney Yang and Yin deficiency syndromes was studied based on decision tree analysis in combination with logistic regression analysis.

It is worthy of study of the relationship between KDS and its symptoms in TCM. According to basic theory of TCM, the kidneys govern reproduction and store essence that serves as the source of marrow, blood and brain. Kidney plays an important role in human growth, development, strong and aging process [27]. Lots of researches confirmed many chronic diseases are closely related to the kidney [28, 29]. There is a saying in TCM: "A chronic disease will inevitably reach the kidneys". The manifestations of KDS can be easily found in 5 chronic kidney related diseases, including osteoporosis, male infertility, female infertility, thalassemia and Alzheimer's disease, so they were studied as our objects. What's more important is that using the symptoms of five diseases as the base factors to be screened can effectively avoid the deviation that a symptom common in one specific disease being selected as a significant symptom related to KDS.

When kidney-Yang is deficient, the fire of the gate of vitality fails to warm the body, causing the aversion to cold as well as the soreness of the back and weakness of the legs and knees. Thus results in impotence and premature ejaculation for men and infertility or frigidity for women [30]. As for kidney Yang deficiency, Qi fails to transform the fluids, therefore resulting in frequent and increased urination as the fluids accumulating. The reasons of KDS-YANG include: a chronic illness after a protracted period of time, excessive sexual activity, constitutional kidney Yang deficiency and decline of the fire of the gate of vitality in the aged. The kidney Yin deficiency causes the arousal of empty-heat within the kidneys, afterward the 5-palm heat, night sweating, red tongue and rapid pulse appears. Yin being deficient and failing to hold defensive Qi in the body at night may cause night sweating. As for

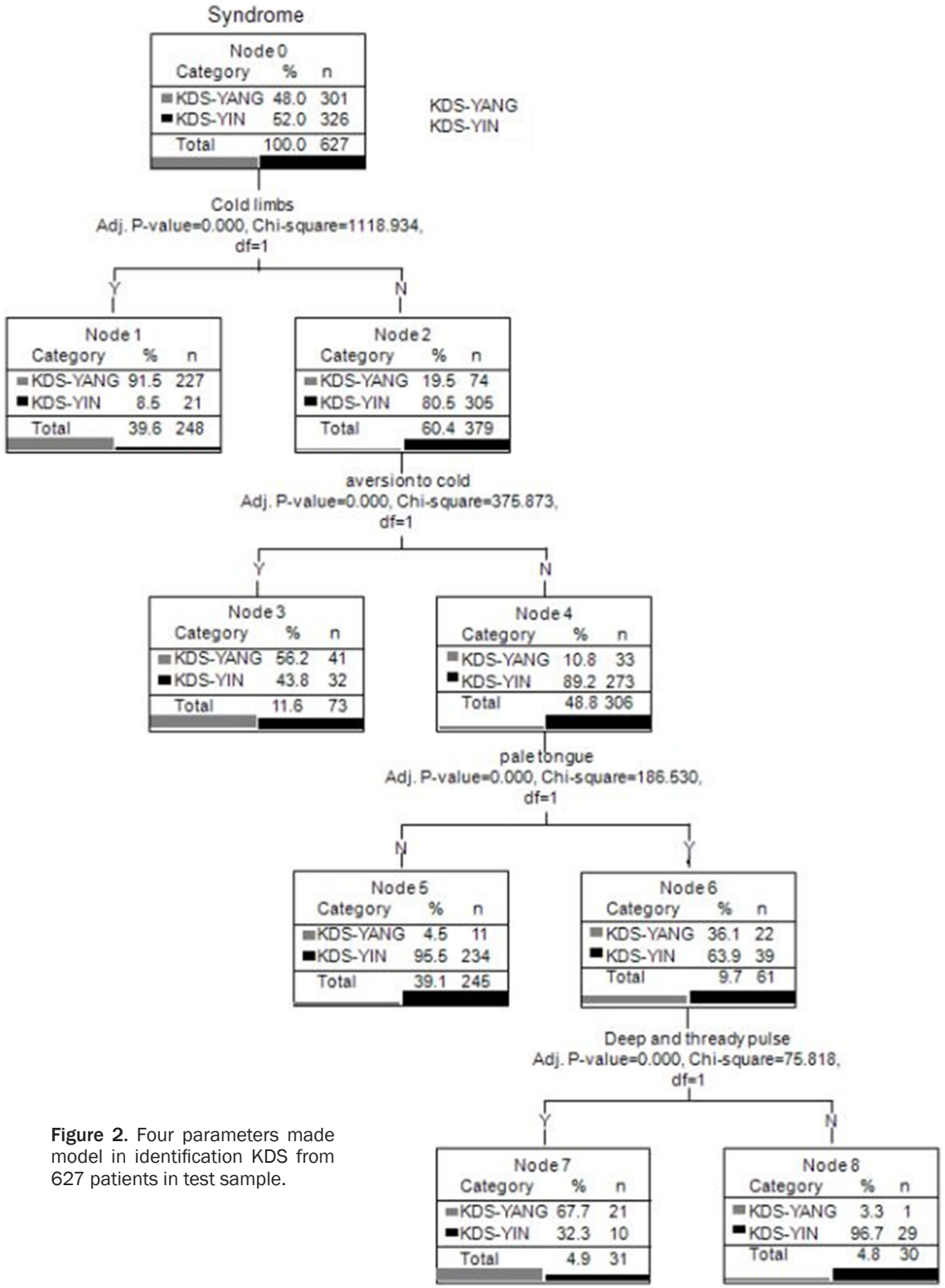

Figure 2. Four parameters made model in identification KDS from 627 patients in test sample.

kidney Yin deficiency, deficiency of essence brings about nocturnal emissions. thus results in the sore back and ache in the bones [31]. The reasons of kidney Yin deficiency include: a long and chronic illness usually transmits from the liver, heart or lungs; overwork over a period

of several years and excessive sexual activity depletes kidney-essence; and it is actually said in TCM that "liver and kidney share the same root". Accordingly loss of blood over a long period of time can cause deficiency of liver blood, which, in turn, can lead to KDS-Yin.

The innovation of the research is that the two core symptoms of differentiating between KDS-YANG and KDS-YIN are obtained based on decision tree analysis in combination with logistic regression. The sensitivity and specificity of the method was achieved satisfactory results so as to assist clinicians to diagnose 5 kidney related diseases. Decision tree was a decision-supporting tool that used for some box-shaped regions or model of decisions based on a logic function, which was comprised of nodes and binary splits [32]. With the development of algorithm of decision tree, it was successfully applied to many fields of syndrome identification in TCM, for example, researches on gastritis [33]. KDS is one of common syndromes in 5 chronic kidney related diseases. So it is not only necessary but also possible to construct KDS model based on combination of logistic regression analysis and QUEST algorithm analysis.

Using the QUEST decision tree, an identification model between KDS-YANG and KDS-YIN was established with four significant variables: two main symptoms, one tongue sign and one pulse sign in the study. Cases with KDS-YANG and KDS-YIN were diagnosed by 8 paths. The above-mentioned four significant variables, namely, aversion to cold, cold limbs, a pale tongue, and a deep and thready pulse were sorted out by the QUEST algorithm decision tree, which were regarded as the decision tree root. The four variables with the most statistical significance would be split if the *P* value was less than α value which was set at 0.05; otherwise the node was not split and was then considered as a terminal node. If the value of the variables such as aversion to cold, cold limbs, deep and thready pulse was set as "yes", it would enter the branch of KDS-YANG; otherwise the node would be split and the variable would continually be used to make the two-dimensional classifications. If the value of the variables such as a pale tongue was set as "none", it enters the branch of KDS-YIN; otherwise the node was split and the variable would continually be used to classify the variables based on the best two dimensional classification tables. Process would be repeated until the *P* value was greater than α value which was set at 0.05, then the classification stopped and finally the model was formed. The results showed that symptoms as aversion to cold, cold limbs, a pale tongue, and a deep and thready pulse were the significant variables to differentiate between KDS-YANG and KDS-YIN.

Cases on five diseases related to KDS had complex clinical manifestations, so it was the basis of syndrome differentiation in TCM. The study showed that identification model was initially constructed to differentiate between KDS-YIN and KDS-YANG based on logistic regression analysis and the QUEST decision tree. The results showed that the above-mentioned 14 variables had the most obvious statistical significance ($P < 0.05$) to distinguish difference between KDS-YANG and KDS-YIN based on logistic regression analysis. Only four variables of aversion to cold, cold limbs, a pale tongue and a deep and thready pulse were selected out among the QUEST decision tree model. As illustrated in **Figure 1**, the results showed that a combination of the four variables was significant differences ($P < 0.05$) in differentiation between KDS-YANG and KDS-YIN based on QUEST decision tree while the decision tree model had diagnostic value. In the study, the diagnostic accuracy of the two analysis method was basically identical. The diagnostic accuracy of logistic regression model was 94.5%, and the identification accuracy of the QUEST decision tree was 88.0%. Accuracy of identification model was ideal because it was consistent with the theory of TCM. However, Logistic regression models used 14 variables, while the decision tree model used only 4 variables concisely.

In summary, symptom of TCM was one of the great helpful tools to diagnose disease. The relevance between TCM syndrome and clinical indicator was explored by using decision tree method in order to make up for setback of the traditional statistical methods and establish the path to objective syndrome diagnosis [34]. The major attributes for diagnosis of KDS-YIN and KDS-YANG were extracted from the complicated data through multi-center clinical epidemiology investigation by using decision tree. Therefore an intuitively clear combi-

nation of identification model was formed, which played a decisive role in the diagnosis of KDS. Decision tree method showed a certain advantage in the objective study on syndrome diagnosis, so it was worth exploring in the field of KDS and the objective study on TCM. So a decision tree analysis model was an effective approach to differentiate KDS-YIN and KDS-YANG, and it could be helpful for shifting the way of syndromes diagnosis from the experience-based form to the data-model based form.

Conclusion

Decision tree model contributed a lot to the identification of KDS patients. In our study, KDS was the basic syndrome of patients related to five diseases: osteoporosis, male infertility, thalassemia, female infertility and Alzheimer's disease. Moreover, the syndrome identification models between KDS-YIN and KDS-YANG was established based on logistic regression analysis in combination with QUEST decision tree analysis. Results showed KDS identification model included four major predictor symptoms: aversion to cold, cold limbs, a pale tongue and a deep and thready pulse. The accuracy of the model was 88.0%, the sensitivity was 96.0% and specificity was 80.7%. A decision tree model was an effective approach to differentiate between KDS-YIN and KDS-YANG, which could be beneficial for shifting the way of diagnosing syndrome from experience-based form to the data-model based form. Constructions of the kidney deficiency validation models would improve further reliability of the identification models.

Acknowledgements

This study was sponsored by the National Key Basic Research and Development Program (973 Program, ID: 2010CB530401).

Disclosure of conflict of interest

None.

Authors' contribution

Tieniu Zhao drafted the manuscript, and Jingyan Meng contributed to the design of the study and modified the manuscript. All authors contributed to the creation of the manual of procedures and acquisition of data. All authors provided critical revision and approved the final manuscript.

Address correspondence to: Jingyan Meng, College of Traditional Chinese Medicine, Tianjin University of Traditional Chinese Medicine, Tianjin 300193, PR China. E-mail: mengjy16@126.com

References

[1] Wang J, Wang PQ, Xiong XJ. Current situation and perspectives of clinical study in integrative medicine in China. Evid Based Complement Alternat Med 2012; 2012: 268542.

[2] Li S, Zhang ZQ, Wu LJ, Zhang XG, Li YD, Wang YY. Understanding ZHENG in traditional Chinese medicine in the context of neuro-endocrine-immune network. IET Syst Biol 2007; 1: 51-60.

[3] Lu A, Jiang M, Zhang C, Chan K. An integrative approach of linking traditional Chinese medicine pattern classification and biomedicine diagnosis. J Ethnopharmacol 2012; 141: 549-556.

[4] Chen KJ, Li L. Study of traditional Chinese medicine-which is after all the right way? Chin J Integr Med 2005; 11: 241-242.

[5] Giovanni Maciocia. The Foundations of Chinese Medicine, British Library Cataloguing in Publication Data; 2005. pp. 417-418.

[6] Wang LF. Diagnostics of Traditional Chinese Medicine, Shanghai: Publishing House of Shanghai University of Traditional Chinese Medicine. 2002; 231-235.

[7] Yang F, Tang DZ, Cui XJ, Holz JD, Bian Q, Shi Q, Wang YJ. Classic yin and yang tonic formula for osteopenia: study protocol for a randomized controlled trial. Trials 2011; 12: 187.

[8] Wu DX, et al. Textbooks for General Tertiary Education of Chinese Medicine: Principles of Chinese Medicine, Shanghai Scientific and Technical Publishers; 2002.

[9] Xiong X, Chu F, Li H, He Q. Clinical application of the TCM classic formulae for treating chronic bronchitis. J Tradit Chinese Med 2011; 31: 69-72.

[10] Zhu WF, Fei SF, Yang MQ, et al. Textbooks for General Tertiary Education of Chinese Medicine: Diagnosis of Chinese Medicine. Shanghai: Shanghai Scientific and Technical Publishers; 1994.

[11] Zhang GG, Singh B, Lee W, Handwerger B, Lao L, Berman B. Improvement of agreement in TCM diagnosis among TCM practitioners for persons with the conventional diagnosis of rheumatoid arthritis: effect of training. J Altern Complement Med 2008; 14: 381-386.

[12] Davis SR, Briganti EM, Chen RQ, Dalais FS, Bailey M, Burger HG. The effects of Chinese medicinal herbs on postmenopausal vasomotor symptoms of Australian women. A randomised controlled trial. Med J Aust 2001; 174: 68-71.
[13] Wang Y, Ma LZ, Liao XW, Liu P. Decision tree method to extract syndrome differentiation rules of posthepatitic cirrhosis in traditional Chinese medicine. IEEE International Symposium on IT in Medicine and Education: 12-14 December 2008; Xiamen. 2008, New York: IEEE; 2008. pp. 744-748.
[14] Shi Q, Zhao H, Chen J, Ma X, Yang Y, Zheng C, Wang W. Study on TCM Syndrome Identification Modes of Coronary Heart Disease Based on Data Mining. Evid Based Complement Alternat Med 2012; 2012: 697028.
[15] Chen XY, Ma LZ, Chu N, Zhou M, Hu Y. Classification and Progression Based on CFS-GA and C5.0 Boost Decision Tree of TCM Zheng in Chronic Hepatitis B. Evid Based Complement Alternat Med 2013; 2013: 695937.
[16] Chen RQ, Wong CM, Lam TH. Construction of a traditional Chinese medicine syndrome-specific outcome measure: the Kidney Deficiency Syndrome Questionnaire (KDSQ). BMC Complement Altern Med 2012; 12: 73.
[17] Liu ZH, Duan YB, Ma HB, et al. Identification and screening of osteoporosis. Chinese J Geront 1995; 15: 53-57.
[18] In: Zheng XY, editor. The Guiding Principles of Clinical Research of New Chinese Medicine (trial). Beijing: China Pharmaceutical Technology Publishing House; 2002.
[19] Shen ZY, Wang WJ. Reference standard on differentiation of deficiency syndrome of traditional Chinese medicine. Chinese J Integr Med 1986; 6: 598.
[20] In: Le J, editor. Obstetrics and gynecology. 7th edition. Beijing: People's Health Press; 2008. pp. 351-352.
[21] World Health Organization (WHO). Standard Inspection Manual of Infertile Couples. Beijing: Peking Union Medical College Hospital, National Family Planning Commission, Institute of Science and Technology; 1995. pp. 44.
[22] In: Zhang ZN, editor. Diagnosis and Efficacy Standard of Blood Disease, Beijing: Science Press; 1991. pp. 48-59.
[23] McKhann G, Drachman D, Folstein M, Katzman R, Price D, Stadlan EM. Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 1984; 34: 939-44.
[24] Identification and Statistical Manual of Mental Disorders. 4th edition. Washington, DC: American Psychiatric Association; 1990.
[25] International Classification of Diseases; 10th revision. Geneva, Switzerland: World Health Organization; 1992.
[26] Kass GV. An exploratory technique for investigating large quantities for categorical data. App Stat 1980; 20: 119-127.
[27] Wei M, Zhao XS, Sun XM, et al. Comparative study on gene expression of kidney-Yin deficiency and kidney-Yang deficiency. J Shanxi Medl Uni 2012; 43: 652-654.
[28] Dong Y, Ding Y, Liu PZ, Song HY, Zhao YP, Li M, Shi JR. Investigation of the Material Basis Underlying the Correlation between Presbycusis and Kidney Deficiency in Traditional Chinese Medicine via GC/MS Metabolomics. Evid Based Complement Alternat Med 2013; 2013: 762092.
[29] Hao YM, Hong MC, Wang WJ. Study on proteins in urine of chronic renal failure patients of different TCM syndrome types. J Integr Tradit Western Med 2012; 32: 1196-1199.
[30] Lian F, Jiang XY, Sun ZG, et al. Study on gene expression profiles in Ovarian Granulosa cells of infertility patients with Kidney-Yin or Kidney-Yang Deficiency Syndrome. J Tradit Chinese Med 2015; 56: 143-147.
[31] Li SQ, Feng EY, Zhang YY, et al. Study of the gene expression profile in the bone tissue with kidney yin deficiency syndromes in primary osteoporosis. Chinese J Osteoporos 2013; 19: 1215-1218.
[32] Neumann A, Holstein J, Le Gall JR, Lepage E. Measuring performance in health care: case-mix adjustment by boosted decision trees. ArtifIntell Med 2004; 32: 97-113.
[33] Zhong Y, Hu XL, Lu JF. Diagnosis in traditional Chinese medicine on gastritis based on relation rules and decision trees. Chinese J Inform Tradit Chinese Med 2008; 15: 97-99.
[34] Shi Q, Chen JX, Zhao HH, et al. Establishment of Qi deficiency syndrome and physicochemical index association patterns in coronary heart disease with diabetes patients based on decision tree. China J Tradit Chinese Med Pharm 2012; 27: 1538-1540.