Contents lists available at ScienceDirect

Digital Chinese Medicine

journal homepage: http://www.keaipublishing.com/dcmed

Grey correlation analysis on influencing factors of Yang deficiency constitution

LUO Yue[a], JIANG Luxia[a], YANG Shengwen[b, c], SU Biliang[b, d], OU Jintao[e], WEN Chuanbiao[a*]

a. College of Intelligent Medicine, Chengdu University of Traditional Chinese Medcine, Chengdu, Sichuan 611137, China

b. Department of Traditional Chinese Medicine, Xi'an Road Community Health Service Center, Jinniu District, Chengdu, Sichuan 610032, China

c. Department of Orthopaedic, Jinniu District Hospital of Traditional Chinese Medicine, Chengdu, Sichuan 610083, China

d. Department of Rehabilitation, Tianhui Community Health Service Center, Jinniu District, Chengdu, Sichuan 610083, China

e. Department of Chinese Internal Medicine, Sichuan Yilong County Hospital of Traditional Chinese Medicine, Nanchong, Sichuan 637676, China

ARTICLE INFO

Article history
Received 20 March 2023
Accepted 24 May 2023
Available online 25 June 2023

Keywords
Yang deficiency constitution
Grey correlation
Influencing factor
Random forest
Model

ABSTRACT

Objective To explore the influencing factors of Yang deficiency constitution in traditional Chinese medicine (TCM) from the perspective of mathematics with the use of calculation formulas, so as to protect patients from getting diseases caused by Yang deficiency constitution and provide suggestions for TCM disease prevention.

Methods Based on the classification and determination criteria of TCM constitution implemented by China Association of Chinese Medicine, data for 24 solar terms from May 5, 2020 (Start of Summer) to April 20, 2021 (Grain Rain) for the identification of Yang deficiency were collected by mobile constitution identification system. The grey correlation analysis method was used to determine the grey correlation degree of the factors influencing Yang deficiency constitution. In addition, a random forest model was constructed for the verification of the results from the grey correlation analysis, and for the evaluation of correlation degree between Yang deficiency constitution and its influencing factors.

Results A total of 16 259 sets of data were collected from healthy or sub-healthy individuals aged from 18 to 60 years living in the central and northeastern parts of Sichuan Province (China) for the identification of TCM constitutions. After screening and preprocessing, a total of 544 sets of data for the identification of Yang deficiency constitution, involving 18 aspects of factors influencing Yang deficiency constitution. The results of the grey correlation analysis showed that there were 12 influencing factors whose grey correlation degree with Yang deficiency constitution was greater than 0.6. The accuracy of these 12 influencing factors with the training set and validation set of the Yang deficiency constitution random forest model were 98.39% and 93.12%, respectively.

Conclusion In the sample data selected in this paper, grey correlation analysis is the appropriate technology to analyze the influencing factors of Yang deficiency constitution. It provides a new idea and a new methodological reference for the research and analysis of the influencing factors of TCM constitution.

*Corresponding author: WEN Chuanbiao, E-mail: wcb@cdutcm.edu.cn.

Peer review under the responsibility of Hunan University of Chinese Medicine.

DOI: 10.1016/j.dcmed.2023.07.005

Citation: LUO Y, JIANG LX, YANG SW, et al. Grey correlation analysis on influencing factors of Yang deficiency constitution. Digital Chinese Medicine, 2023, 6(2): 151-159.

1 Introduction

Traditional Chinese medicine (TCM) constitutions are crucial for “preventive treatment of disease” in TCM. “Saint does not treat diseases, but prevents the disease from occurring” is a typical TCM thinking that comes from the TCM classic *Plain Question*: *The Great Theory of Four Qi Regulating Spirits* (*Su Wen · Si Qi Tiao Shen Da Lun Pian*, 《素问·四气调神大论篇》). The idea of "preventive treatment of disease" is of great significance to the practice of healthy China strategy [1]. Constitution refers to the comprehensive and relatively stable inherent characteristics of morphological structure, physiological function, and psychological state that are shaped throughout a person’s life, influenced by both innate attributes and acquired factors. It is the unique characteristics of the human body that develop and adapt in harmony with the natural and social environment throughout the process of human growth and development [2]. WANG Qi, a renowned expert in TCM, classified constitutions into nine types: Balanced constitution, Qi deficiency constitution, Yang deficiency constitution, Yin deficiency constitution, Phlegm dampness constitution, Damp heat constitution, Hemostasis constitution, Qi stagnation constitution, and Inherited Special constitution [2]. Except for the Balanced constitution, the other eight constitutions are all imbalanced types, characterized by their susceptibility to specific causes and diseases, as well as certain tendencies in the progression and outcome of diseases [2]. They are deemed the “common soil” for many diseases. The Yang deficiency constitution is primarily characterized by a deficiency of Yang Qi, an aversion to cold, and a constitution marked by cold and deficient manifestations [3-5]. WANG Qi highlighted that individuals with Yang deficiency constitution were more susceptible to diseases such as arthritis, cough, asthma, chronic enteritis, and dysmenorrhea [5], and numerous studies have corroborated this perspective [6-10]. Indeed, some studies have indicated that Yang deficiency constitution is associated with an increased risk of cardiovascular disease, urticaria, metabolic syndrome, cancer, sepsis, and diabetes in the elderly [11-19]. Constitution is a relatively stable state of the human body, yet it also possesses dynamic adjustability. Therefore, studying the influencing factors of Yang deficiency constitution is beneficial for proactively preventing related diseases, thereby achieving the goals of “preventive treatment of disease” and “prevention before disease onset”.

The grey correlation analysis is an essential component of the grey analysis theory, which was proposed by Professor DENG Julong in China [20]. It is a novel approach used to study problems that have limited information and uncertainty. It is a factor analysis method that leverages limited data and partial known information to extract valuable insights. The grey correlation analysis helps analyze and determine the degree of influences between factors or their contribution measures to the main features under study, which is especially useful when dealing with small samples or situations where complete data is unavailable, making it a powerful tool in decision-making and problem-solving process [21]. It is an approach that has been employed in various research fields [22-24], including the field of TCM research. Based on the grey correlation theory, TIAN et al. [25] analyzed the correlation between the identification of “Jing-Qi-Shen” and each sub-identification system of the health status of diabetic patients in TCM diagnosis. Their conclusion was that when the identification of “Jing-Qi-Shen” was insufficient in diabetic patients, the most relevant factor was insufficient “Jing”, followed by insufficient “Qi” “Qi loss” and finally "Shen loss". Compared with methods such as logistic regression analysis, analysis of variance, analytic hierarchy process, and descriptive statistics, the grey correlation analysis method has no requirements for sample size, without typical distribution rules that the data need to follow. The calculation of the grey correlation analysis is simple, with results obtained more similar to the qualitative analysis results. The TCM constitutions collected in this study encompassed the results of the scales as well as their respective constitution types filled out by individuals themselves. If an individual completes the collection independently without the guidance of a doctor, the collection process may be uncontrollable, resulting in unavoidable information gaps and incompleteness in the collected TCM constitution data. Therefore, taking into account the aforementioned reasons, this study uses the grey correlation analysis method to explore the influencing factors of Yang deficiency constitution in TCM, in order to offer new insights and methodological references for the analysis of influencing factors of TCM constitution.

2 Data and methods

Data from the central and northeastern parts of Sichuan Province (China) of the identification of TCM constitutions (ethical approval number: 2022KL-024, approved by the Medical Ethics Committee of the Affiliated Hospital of Chengdu University of Traditional Chinese Medicine) were collected from May 5, 2020 (Start of Summer) to April 20, 2021 (Grain Rain) by using the mobile acquisition instrument developed for the specific purpose. All subjects filled out the TCM constitution identification scale after submitting their informed consent.

2.1 Inclusion criteria

(i) Patients were eligible if they were aged between 18 and 60 years (without gender bias); (ii) they were either

healthy or sub-healthy; (iii) they met the TCM Constitution Classification and Determination Criteria [3] implemented by the China Association of Chinese Medicine for Yang deficiency constitution, with a conversion score ⩾ 40; (iv) they were informed and consent of the study, and voluntarily submitted their relevant information.

2.2 Exclusion criteria

(i) Patients were excluded if they were in the treatment process of cardiovascular and cerebrovascular diseases, diabetes, tumor, and other major diseases; (ii) they suffered from non-major diseases but medication maintenance was required; (iii) they were pregnant or lactating; (iv) they had mental or neurological diseases; (v) their constitution identification data was combined with Yang deficiency constitution.

2.3 Observed indicators

A total of 19 variables including gender, age, birth mode, feeding mode within four months after birth, age of father at birth, age of mother at birth, time of birth, family history of diseases, sleeping habits, exercising habits, major changes in work or life, eating habits, solar terms, living environment, menstrual period (only for women), overtime work, body mass index (BMI), sleeping duration, and TCM constitution type, were considered. Among the 19 variables, the former 18 were treated as independent ones, while the last one was dependent variable.

2.4 Data analysis

2.4.1 Data quantification In the grey correlation analysis, numerical data are key for calculations, so it is necessary to identify the numerical data of TCM constitution identification (constitution type and its influencing factor). In this study, a total of nine TCM constitutions were identified, which were Balanced constitution, Qi deficiency constitution, Yang deficiency constitution, Yin deficiency constitution, Phlegm dampness constitution, Damp heat constitution, Hemostasis constitution, Qi stagnation constitution, and Inherited Special constitution, and were marked by the numbers 1 to 9, respectively.

An $18 * n$ matrix was set to quantify the influencing factors of 18 constitutions, which were numbered from 1 to 18, with the integer numbers used in a sequential order. Taking an individual with Yang deficiency constitution as an example, the specific conditions of the individual were: female, 18 years old, born at time of Zi Shi (from 11 p.m. to 1 a.m in a day), non-menstrual period, no family history of diseases, not very active, Grain Rain solar term, spontaneous birth at birth, mother being 24 years old and father being 26 at the time of her birth, breastfeeding within four months after birth, irregular sleep, 7 h for sleeping per day, no overtime work, no dead relatives, a spicy food lover, living environment in a suitable environment, BMI 20. The influencing factors of the constitution of this individual were quantified (Table 1).

Table 1 A quantitative table of the influencing factors for an individual with Yang deficiency constitution

No.	Category of entries for selection	Individual reality	Quantified entry
1	Gender (male, female)	Female	2
2	Age	19 years old	19
3	Time of birth (Zi Shi, Chou Shi, Yin Shi, Mao Shi, Chen Shi, Si Shi, Wu Shi, Wei Shi, Shen Shi, You Shi, Xu Shi, Hai Shi)	Zi Shi	1
4	Menstrual period (no, yes, none)	No	1
5	Family history of diseases (hypertension, diabetes mellitus, gastrointestinal diseases, heart diseases, mental diseases, hyperlipidemia, liver diseases, malignant tumors, stroke, allergic diseases, others, none)	None	12
6	Exercising habits (Exercise often, Sometimes exercise, Do not exercise much)	Do not exercise much	3
7	Solar term (Beginning of Spring, The rains, the Waking of Insects, Spring Equinox, Pure Brightness, Grain Rain, Beginning of Summer, Grain Buds, Grain in Ear, Summer Solstice, Minor Heat, Major Heat, Beginning of Autumn, End of Heat, White Dew, Autumn Equinox, Cold Dew, First Frost, Beginning of Winter, Minor Snow, Major Snow, Winter Solstice, Minor Cold, Major Cold)	Grain Rain	6
8	Birth mode (normal delivery, cesarean section, premature delivery, difficult delivery, unknown)	Normal delivery	1
9	Age of mother at birth	24 years old	24
10	Age of father at birth	26 years old	26
11	Feeding mode within four months after birth (breast feeding, mixed feeding, artificial feeding, unknown)	Breast feeding	1
12	Sleeping habits (go to bed and get up early, go to bed early but get up late, go to bed late but get up early, go to bed and get up both late)	Irregular	5

Table 1 Continued

No.	Category of entries for selection	Individual reality	Quantified entry
13	Sleeping duration (hours per day)	7	7
14	Overtime work (hours per week) (⩾20, 10 - 20, 5 - 10, <5, no overtime)	No overtime work	5
15	Major changes in work or life (death of a loved one, divorce, property loss, unemployment, none)	Death of a loved one	1
16	Eating habits (prefer sweet, prefer spicy, prefer sour, prefer salty, prefer light, prefer greasy, prefer roasted, prefer cold, prefer cool, prefer hot, good at smoking, good at drinking alcohol, good at drinking tea, others, no particulars)	Prefer spicy	2
17	Living environment (dark and damp, dry, suitable environment)	Suitable	3
18	BMI (kg/m^2)	50 kg, 1.58 m	20

Then the quantization matrix of the body identification data for this individual with Yang deficiency constitution was:

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 & 13 & 14 & 15 & 16 & 17 & 18 \\ 2 & 19 & 1 & 1 & 12 & 3 & 6 & 1 & 24 & 26 & 1 & 5 & 7 & 5 & 1 & 2 & 3 & 20 \end{bmatrix}.$$

2.4.2 Analysis process Based on the grey correlation analysis method, the analysis of the influencing factors of TCM constitution includes the following steps.

(i) Determination of the reference and comparison sequences in the data matrix for TCM constitution recognition. The sequence comprising the dependent variables, which was the Yang deficiency constitution, was expressed as the reference sequence $\{Y_i\}$. The sequence comprising the independent variables, the influencing factors of the TCM constitution, was denoted as the comparison sequence $\{X_i\}$. Each sequence was established with the use of TCM body identification information collected from the individuals at different time points. For example, for Y_i, the TCM constitution type of the first individual was denoted as y_1, the second individual as y_2, the ith individual as y_i. Similarly, the comparison sequence was $X_i = [x_i(1),\ x_i(2)...x_i(n)]$, $x_i(1)$ indicated the first influencing factor for the first individual, $x_i(2)$ indicated the second influencing factor of the second individual, $x_i(n)$ indicated the nth influencing factor of the nth individual.

(ii) Dimensionless data processing based on mean value method. The reference and comparison sequences were subjected to dimensionless processing, leading to a dimensionless data matrix. Dimensionless methods usually encompass the initial value method and the mean value method. In the experiments, it was observed that the application of the initial value method resulted in more volatility and inaccuracies in the experimental results. To mitigate this situation, the mean value method was selected for the dimensionless processing of the data matrix to enhance the accuracy of the experimental results.

Let $X_i = [x_i(1),\ x_i(2)...x_i(n)]$ be a comparison sequence of influencing factors X_i, then the sequence was dimensionlessly processed by the mean method as:

$$x_i(k) = \frac{x_i(k)}{\overline{X_i}},\ \overline{X_i} = \frac{1}{n}\sum_{k=1}^{n} x_i(k),\ k = 1,2,...,n \tag{1}$$

$\overline{X_i}$ in the formula was the mean value of the ith comparison sequence. The Yang deficiency constitution identification data (reference and comparison sequences) were processed dimensionlessly using Equation (1).

(iii) Calculations of the difference sequence, as well as the maximum and minimum differences between the comparison sequence and the reference sequence for the Yang deficiency constitution type.

Calculating the difference sequence:

$$\Delta i(n) = |Y_i - X_i(n)| \tag{2}$$

Calculating the minimum difference:

$$\Delta min = \min_i \min_n \Delta i(n) = \min_i \min_n |Y_i - X_i(n)| \tag{3}$$

Calculating the maximum difference:

$$\Delta max = \max_i \max_n \Delta i(k) = \max_i \max_n |Y_i - X_i(n)| \tag{4}$$

In Equation (2), $\Delta i(n)$ was the absolute value of the difference sequence between the comparison sequence and the reference sequence. In Equation (3), $\min_n \Delta \mathrm{i}(n)$ was the minimum difference sequence of the nth comparison sequences for the ith individual. On this basis, $\min_i \min_n \Delta i(n)$ was the absolute value of the minimum difference between the comparison sequence and the reference sequence for the ith individual. Equation (4) was the same as Equation (3) for the maximum difference sequence. The difference sequence, maximum difference sequence, and minimum difference sequence of the dimensionless data were found respectively with the use of Equations (2), (3), and (4).

(iv) Calculation of the grey correlation coefficients for each influencing factor of Yang deficiency constitution.

$$\S_i(k)=\frac{\min\limits_i\min\limits_k|Y_i-X_i(n)|+\xi\max\limits_i\max\limits_k|Y_i-X_i(n)|}{|Y_i-X_i(n)|+\xi\max\limits_i\max\limits_k|Y_i-X_i(n)|} \tag{5}$$

In Equation (5), $\S_i(k)$ was the grey correlation coefficient for each influencing factor. ξ was the resolution coefficient, and its value range was (0, 1). The value of ξ was not explicitly defined; however, in this study, the values were calculated based on the formula derived from the research of ZHANG Ruiqiang [26], WEI Meng [27], and other studies on the resolution coefficient.

$\Delta\upsilon$ was denoted as the mean value of all differences, which is $\Delta\upsilon=\frac{1}{n\cdot m}\sum_{i=1}^{m}\sum_{k=1}^{n}|x_0(n)-x_i(n)|$. The quotient of the mean $\Delta\upsilon$ of all differences was noted with the maximum difference Δmax, which is $X\Delta=\frac{\Delta\upsilon}{\Delta max}$, Then the value of ξ was expressed as $X\Delta\le\xi\le 2X\Delta$, which satisfies the following condition.

$$\Delta max>3\Delta\upsilon, X\Delta\le\xi\le 1.5X\Delta$$
$$\Delta max\le 3\Delta\upsilon, 1.5X\Delta<\xi\le 2X\Delta$$

Through calculation, it was determined in this paper that all data pertaining to the identification of Yang deficiency constitution Δmax were greater than $3\Delta\upsilon$. Meanwhile, WEI Meng [27] and others suggested that when $\Delta max>3\Delta\upsilon$, the resolution coefficient ξ is generally considered as $1.5X\Delta$. Therefore, in this paper, the value of ξ was $1.5X\Delta$ and the accuracy kept three digits after the effective digit decimal point for calculation. All the data pertaining to the identification of Yang deficiency constitution covered in this study, including the maximum difference absolute value Δmax, the mean differences $\Delta\upsilon$, the quotient of the mean difference to the maximum difference $X\Delta$ ($X\Delta=\frac{\Delta\upsilon}{\Delta max}$) and the resolution coefficient ξ were shown in Table 2.

Table 2 The values of $\Delta max, \Delta\upsilon, X\Delta, \xi$, and resolution coefficient for calculations

Index	Δmax	$\Delta\upsilon$	$X\Delta$	ξ	ξ for calculation
Value	3.518 272	0.314 940	0.089 515	0.134 274	0.134 000

(v) Calculation of the correlation degree between the influencing factors and the Yang deficiency constitution.

Calculating the grey correlation degree:

$$\gamma_i=\frac{1}{n}\sum_{k=1}^{n}\S_i(k) \tag{6}$$

In Equation (6), γ_i represented the grey correlation degree of each influencing factor of the Yang deficiency constitution. Equation (6) was utilized to calculate the correlation degree between the reference and comparison sequences in the data for the identification of Yang deficiency constitution based on the scales.

(vi) Correlation degree ranking. To accurately evaluate the correlation between the comparison sequence and each reference sequence, the correlation degree for each data column was calculated, and subsequently, the correlation was ranked based on its magnitude. This ranking is known as the correlation order. A high degree of correlation indicated a close relationship between the comparison sequence and the reference sequence, and a low degree of correlation suggested otherwise. As a result, it becomes feasible to identify both the major and minor factors of the dependent variable (reference sequence).

(vii) The python programming language was used to carry out the grey correlation calculation of the influencing factors of Yang deficiency constitution, and the Gephi software was used to realize the visualization of the correlation.

2.4.3 Validation analysis To verify and analyze the results obtained from the grey correlation analysis method, this study employed random forest model for the construction of a prediction model. The accurate indexes of the model were then utilized to evaluate the effectiveness of the grey correlation analysis method. The random forest model belongs to a prominent branch of machine learning and is commonly used to deal with the problems of classification. The model still employs the bootstrap resampling technique to randomly and repeatedly draw m samples with replacement from the original training sample set M for the generation of a new training sample set. This new set is then used to train decision trees in the model. After generating N decision trees according to the above steps, a random forest model is created. In essence, the created random forest model is an enhancement of the decision tree algorithm that combines multiple decision trees into a unified model. Each tree is constructed based on independently sampled data, and the randomness in the sampling of modeled features endows it with resistance to overfitting. The parallel modeling of multiple decision trees contributs to its high accuracy.

Based on the established conclusion that factors with a grey correlation degree greater than 0.6 are indicative of strong correlations with the subject [28, 29], this study selected influencing factors with grey correlation degrees greater than 0.7 and 0.6, respectively, from the results of the grey correlation analysis of Yang deficiency constitution. These selected influencing factors were then used to construct a random forest classification model, and the accuracy of the model was subsequently calculated. In model building, the random forest model fitting was performed using a training set to validation set ratio of 7 : 3. At the same time, this paper included an equal number of

copies of the data for the identification of the Balanced constitution for validation analysis. The number of decision trees was set to 100, and the maximum depth of the tree was set to 10. Randomly sampled data with replacements were used to form the training data, while the validation data set was used to test the model.

3 Results

3.1 Grey correlation analysis

In this paper, a total of 16 259 TCM constitution identification data were collected from certain areas of central and northeastern Sichuan Province, of which 544 were carefully screened and pretreated to obtain the data of individuals with Yang deficiency constitution. The grey correlation degree of each influencing factor of Yang deficiency constitution was calculated by grey correlation analysis based on the 544 identified and quantified data, as shown in Table 3.

Table 3 Ranking of the influencing factors and their grey correlation degree

No.	Influencing factor	Degree
1	Sleeping duration	0.855 365
2	Major changes in work or life	0.816 374
3	Living environment	0.801 253
4	Age of mother at birth	0.800 638
5	Overtime work	0.795 966
6	Birth mode	0.792 387
7	Age of father at birth	0.771 777
8	BMI	0.765 551
9	Exercising habits	0.684 649
10	Gender	0.641 390
11	Age	0.607 687
12	Feeding mode within four months after birth	0.603 698
13	Eating habits	0.556 708
14	Solar term	0.543 349
15	Time of birth	0.532 215
16	Menstrual period	0.506 412
17	Sleeping habits	0.461 842
18	Family history of diseases	0.434 322

The influencing factors of the Yang deficiency constitution were represented in a bubble diagram, as shown in Figure 1.

Table 3 and Figure 1 showed that there were four influencing factors of Yang deficiency constitution with grey correlation degrees of 0.8 or above (accounting for 22.2% of the total factor numbers). They were sleeping duration, major changes in work or life, living environment, and age of mother at birth. There were four influencing factors whose degree was between 0.7 and 0.8 (22.2%). They are overtime work, birth mode, age of father birth and BMI. There were four whose degree was between 0.6 and 0.7 (22.2%). They were exercising habits, gender, age, feeding mode within four months after birth. There were six factors whose degree was between 0.4 and 0.6 (33.4%). They were eating habits, solar term, time of birth, menstrual period, sleeping habits, and family history of diseases. Based on the results of previous studies [28, 29], the value of the grey correlation degree indicated the strength of the association. Degree between [1, 0.8] represented the strongest association, [0.8, 0.6] a relatively strong association, [0.6, 0.4] a moderately strong association, and below 0.4 a weak association. Therefore, sleeping duration, major changes in work or life, living environment, and age of mother at birth were strongly correlated with Yang deficiency constitution. Overtime work, birth mode, age of father at birth, BMI, exercising habits, gender, age, and feeding mode within four months after birth had a relatively strong correlation with Yang deficiency constitution. Eating habits, solar term, time of birth, menstrual period, sleeping habits, and family history of diseases were moderately associated with Yang deficiency constitution.

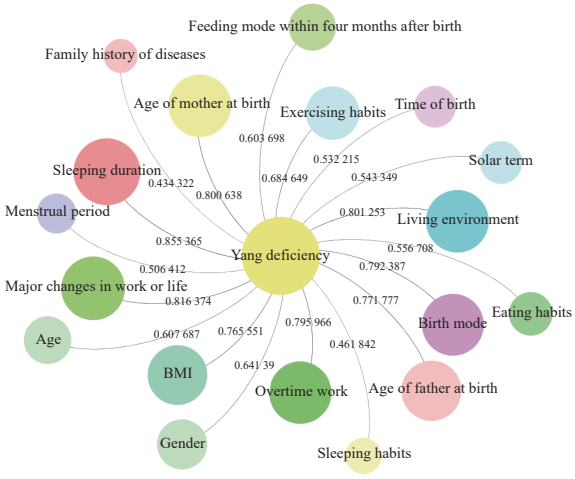

Figure 1 Grey correlation bubble diagram of factors affecting Yang deficiency constitution

Different bubble colors represent different factors, with larger bubbles indicating a greater grey correlations between the factor and the Yang deficiency constitution, and smaller bubbles indicating otherwise.

3.2 Validation of analysis results

A strong correlation was indicative when the grey correlation degree fell to [1, 0.6]. Thus, we constructed the random forest model with the influencing factors whose grey correlation degrees were greater than 0.7 and 0.6. The accuracy of the model was shown in Table 4. The average prediction accuracy of the random forest model constructed using eight influencing factors with grey correlation degrees greater than 0.7 was 94.83% and 85.78% in the training and validation sets, respectively. The average

prediction accuracy of the random forest model constructed using 12 influencing factors with grey correlation degrees greater than 0.6 were 98.39% and 93.12% in the training and validation sets, respectively. AUC stands for area under the receiver operating characteristic (ROC) curve, which is indicative of the prediction accuracy. The AUC value ranges from 0 to 1, and higher value indicates higher accuracy. As shown in Figure 2, in this study, the AUC value of the influencing factors with a correlation degree greater than 0.7 was 0.993 in the random forest training set and 0.953 in the validation set. The influencing factors with correlation degrees greater than 0.6 were 0.999 and 0.988 in the random forest training set and validation set, respectively.

Table 4 The prediction accuracy of the random forest model

Influencing factor	Accuracy	
	Training set	Validation set
Grey correlation degree greater than 0.7	94.83%	85.78%
Grey correlation degree greater than 0.6	98.39%	93.12%

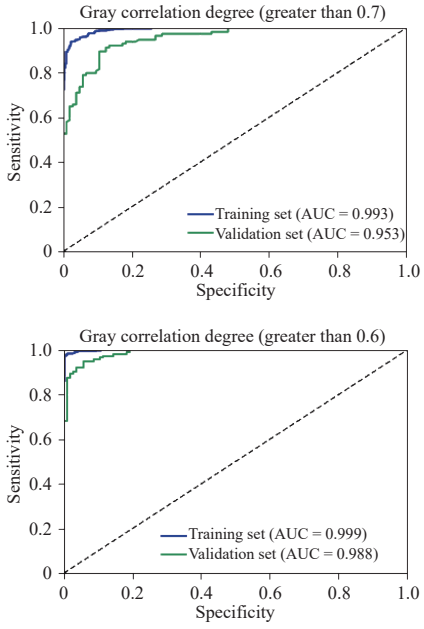

Figure 2 The ROC curve of the influencing factors of Yang deficiency constitution validated by the random forest model

A, with grey correlation degree greater than 0.7. B, with grey correlation degree greater than 0.6.

4 Discussion

Yang deficiency constitution is one of the eight imbalanced constitutions in TCM, characterized by a deficiency of Yang Qi and primarily manifested as a tendency towards coldness, which could cause many diseases. Studying the influencing factors of Yang deficiency constitution can effectively avoid the formation of Yang deficiency constitution, prevent the occurrence of diseases, and achieve the goal of "preventive treatment of disease". After analyzing the connotation of TCM constitutions, WANG Qi [1, 2] proposed that the main characteristics of Yang deficiency constitution included weakness in muscle, lack of energy, excessive sleep, susceptible to dampness, introverted personality, and preference for solitude. At the same time, he highlighted that Yang deficiency constitution could be attributed to various significant factors, including congenital deficiencies, weakened parental health during pregnancy, advanced maternal age, and premature birth. Recent research further supports the notion that residing in an environment with an ideal setting and ample sunshine plays a pivotal role in regulating the Yang deficiency constitution [30]. Due to the deficiency of Yang Qi, individuals with Yang deficiency constitution are not fond of exercising, hence exercising more could improve the imbalanced constitutions as well. Women exhibited a higher susceptibility to the Yang deficiency constitution compared with men, and mitigating this condition could be realized by taking measures to avoid exposures to coldness and other relevant factors [30]. The feeding method during the first four months after birth markedly influences the development of the acquired constitution [31], with breastfeeding proving to have a positive and beneficial impact on infants' nervous system and visual development. Therefore, individuals ought to proactively prevent the development of Yang deficiency constitution by addressing both congenital and acquired factors mentioned above. Additionally, they should focus on regulating this condition through appropriate measures.

Grey correlation analysis offers a distinct advantage over other analytical methods due to its ability to bypass specific requirements concerning the number and regularity of samples. Therefore, in this study, grey correlation analysis was applied to scrutinize the data concerning the identification of Yang deficiency constitution. The analysis revealed 12 influencing factors, including sleeping duration, major changes in work or life, living environment, age of mother at birth, overtime work, birth mode, age of father at birth, BMI, exercising habits, gender, age, and feeding mode within 4 months after birth, exhibiting a robust correlation with Yang deficiency constitution (with grey correlation degrees greater than 0.6). This is basically consistent with the current theoretical and practical research results on Yang deficiency constitution [3, 4, 30, 31].

This is the first study that has explored and analyzed the influencing factors of Yang deficiency constitution with the use of the grey correlation analysis method. The sample data used for the analysis and research are relatively limited in size, and the applicability of the grey correlation analysis to TCM constitutions has not been investigated yet. In the future, it is necessary to carry out

in-depth research on the influencing factors of other imbalanced TCM constitutions with larger sample size.

5 Conclusion

In this study, the grey correlation analysis method is introduced to explore the influencing factors of Yang deficiency constitution from 18 aspects. Subsequently, factors exhibiting grey correlation degrees greater than 0.6 and 0.7 are selected to construct the random forest model. The verification results suggest that the influencing factors with grey correlation degrees greater than 0.6 and 0.7 have a strong correlation with Yang deficiency constitution. Moreover, this analysis affirms grey correlation analysis is the suitable and appropriate technique for examining the influencing factors of Yang deficiency constitution within the selected sample data for this study. This study serves as a valuable through introducing new methods and ideas for the investigation of influencing factors of TCM constitutions. It also provides a novel approach for the clinical treatment and rehabilitation of Yang deficiency constitution in TCM.

Fundings

National Natural Science Foundation (81904324 and 82174236), and Chengdu Science and Technology Bureau Project (2019-YF09-00185-SN).

Competing interests

The authors declare no conflict of interest.

References

[1] FANG YN, WANG Q, ZHANG GH, et al. On application of traditional Chinese medicine constitution in "preventive treatment of disease". Journal of Traditional Chinese Medicine, 2020, 61(7): 581–585.

[2] WANG Q. Constitution of traditional Chinese medicine. Beijing: People's medical Publishing House, 2009.

[3] China Association of Chinese Medicine. Classification and determination of constitution in TCM. Journal of Chinese Society of Traditional Chinese Medicine, 2009, 4(4): 303–304.

[4] WANG Q. Classification and diagnosis basis of nine basic constitutions in Chinese medicine. Journal of Beijing University of Traditional Chinese Medicine, 2005, 28(4): 1–8.

[5] WANG Q, TIAN Y. Decoding the nine types of Chinese constitutions. Beijing: China Traditional Chinese Medicine Press, 2009: 3-23.

[6] ZHOU Q, AN MW, TANG Y, et al. Discussion the relationship between Yang deficiency and cut microbiome. Journal of Jiangxi University of Chinese Medicine, 2023, 35(1): 20–23.

[7] BAI YP, WANG JJ. Research on traditional Chinese medicine constitution and related influencing factors of lumbar disc herniation. Journal of Emergency in Traditional Chinese, 2014, 23(9): 1662–1664.

[8] LI J, LIU Y, CHEN ZX, et al. Investigation of TCM constitution in 2572 female college students with primary dysmenorrhea: a correlation study. Journal of Guangzhou University of Traditional Chinese Medicine, 2022, 39(12): 2752–2761.

[9] SUN CR, ZHANG YL, QI BY, et al. Distribution of traditional Chinese medicine constitution and associated factors of Yang deficiency in patients with postmenopausal osteoporosis in the community. Chinese General Practice, 2022, 25(18): 2201–2206.

[10] HAN J, ZHAO YH, WANG JX, et al. Effects of overweight or obesity on airway inflammation, asthma control and TCM constitution in asthma patients. Modern Journal of Integrated Traditional, 2022, 31(8): 1023–1026, 1033.

[11] ZHAO B, LUO XC, ZHAO X. Analysis of the results of TCM constitution identification in high-risk population with cardiovascular disease. Chinese Journal of Integrative Medicine on Cardio-Cerebrovascular Disease, 2019, 17(6): 801–805.

[12] CHENG QZ, CHEN AL, ZHAO ZK. Analysis of the correlation between traditional Chinese medicine constitution characteristics and BMI, IMT in high-risk population of cardiovascular disease. Chinese Journal of Integrative Medicine on Cardio-Cerebrovascular Disease, 2020, 18(9): 1406–1408.

[13] ZHOU JL, LI SM, QIU HL. Analysis of traditional Chinese medicine constitution distribution and related influencing factors in 150 patients with chronic urticaria. Journal of Guangxi University of Chinese Medicine, 2020, 23(1): 23–25.

[14] XU F, ZHANG YH, CUI WQ, et al. The association between metabolic syndrome and body constitution in traditional Chinese medicine. European Journal of Integrative Medicine, 2017, 14: 32–36.

[15] KUNG YY, KUO TBJ, LAI CT, et al. Disclosure of suboptimal health status through traditional Chinese medicine-based body constitution and pulse patterns. Complementary Therapies in Medicine, 2021, 56: 102607.

[16] TANG SQ, WANG YL, XIE ZY, et al. Serum metabolic profiling of traditional Chinese medicine syndromes in patients with diarrhea-predominant irritable bowel syndrome. Journal of Integrative Medicine-Jim, 2021, 19(3): 274–281.

[17] TIAN XD, ZHU GY, WANG J, et al. Study on the relation between tissues pathologies and traditional Chinese medicine syndromes in knee osteoarthritis: medical image diagnostics by preoperative X-ray and surgical arthroscopy. Journal of X-Ray Science and Technology, 2016, 24(4): 509–519.

[18] LIAO YC, CHEN LL, WANG HC, et al. The association between traditional Chinese medicine body constitution deviation and essential hypertension: a case-control study. Journal of Nursing Research, 2021, 29(4): e160.

[19] ZHANG HX, WANG LM, GUO JP, et al. Gut microbiota and differential genes-maintained homeostasis is key to maintaining health of individuals with Yang deficiency constitution. Journal of Traditional Chinese Medicine, 2022, 42(1): 96–101.

[20] DENG JL. Grey System Theory Tutorial. Wuhan: Huazhong University of Technology Press, 1990.

[21] LIU SF. Emergence and development of grey system theory and its forward trends. Journal of Nanjing University of

Aeronautics & Astronautics, 2004, 49(2): 267–272.
[22] LI WH, XU JH, SUN CG. Spatial and temporal evolution of Shenzhen's eco-environment and urban sprawl based on grey correlation degree. Ecology and Environmental Sciences, 2021, 30(4): 880–888.
[23] GAO Q, CHEN JN, ZHANG JC, et al. Comprehensive evaluation of the effects of different drying methods on the quality of Chinese yam chips based on gray relational analysis. Science and Technology of Food Industry, 2018, 39(16): 6–12.
[24] LIU FC, LIU AL. A study on relationship of tourism factors based on the method of gray relational analysis—a case study of Beijing inbound tourism market. Journal of Capital Normal University (Natural Science Edition), 2016, 37(3): 84–90.
[25] TIAN ZK, WANG DJ, ZHOU ZH, et al. Identification and analysis of "Jing-Qi-Shen" in patients with diabetes based on grey system theory. Lishizhen Medicine and Materia Medica Research, 2021, 32(5): 1267–1269.
[26] ZHANG RQ, QU XZ, JI M. Determination of distinguishing coefficient of gray correlation degree and application in mechanical fault diagnosis. Coal Mine Machinery, 2013, 34(3): 291–293.
[27] WEI M, GAO P, FENG XZ, et al. Application of improved grey correlation method in debris flow hazard assessment. Gansu Water Resources and Hydropower Technology, 2020, 56(10): 16–20.
[28] BAO HM, CHI ET, YE TF. Grey correlation analysis of factors affecting ultraviolet aging of colored asphalt mixture. Journal of Huaqiao University (Natural Science), 2020, 41(4): 466–473.
[29] PAN ZX, XU K. Research on innovation and entrepreneurship policy efficiency based on grey correlation analysis. Journal of Entrepreneurship in Science & Technology, 2021, 34(4): 13–17.
[30] YE ZY, LI H, CHEN XY, et al. Analysis of environmental influencing factors on Yang deficiency constitution. Guangdong Medical Journal, 2017, 38(11): 1659–1662.
[31] LUO H, LI YS, LI ZQ, et al. Correlation between Yang deficiency constitution and disease: based on bibliometric analysis on 214 clinical studies. Journal of Traditional Chinese Medicine, 2020, 61(8): 716–720.

基于灰色关联分析法的阳虚质影响因子分析

罗悦[a], 蒋璐霞[a], 杨生文[b,c], 苏必良[b,d], 欧金涛[e], 温川飙[a*]

a. 成都中医药大学智能医学学院, 四川 成都 611137, 中国
b. 四川省成都市金牛区西安路社区卫生服务中心中医科, 四川 成都 610032, 中国
c. 四川省成都市金牛区中医医院骨科, 四川 成都 610083, 中国
d. 四川省成都市金牛区天回社区卫生服务中心康复科, 四川 成都 610083, 中国
e. 四川省仪陇县中医医院中医内科, 四川 南充 637676, 中国

【摘要】目的 运用数学思维和计算方法，探索中医阳虚质影响因子以避免由阳虚质引起的诸多疾病，为中医治未病提供方案建议。**方法** 基于中华中医药学会发布的中医体质分类与判定标准，利用项目组移动中医体质辨识系统采集 2020 年 5 月 5 日（立夏）到 2021 年 4 月 20 日（谷雨）24 节气阳虚体质辨识数据，运用灰色关联分析法分析得出阳虚质影响因素的灰色关联度。在此基础上构建随机森林模型对灰色关联分析法结果进行验证分析，评价阳虚质与其影响因子关联的强弱。**结果** 本研究采集了来自中国四川省中部和东北部地区年龄在 18-60 岁的健康或亚健康人群，共计 16 259 份中医体质辨识数据。经过筛选、预处理得到阳虚体质辨识数据共 544 份，涉及阳虚质影响因子 18 个侧面。运用灰色关联分析法分析后，灰色关联度大于 0.6 的影响因子有 12 个，这 12 个影响因子与阳虚质随机森林模型训练集和验证集的准确率分别为 98.39% 和 93.12%。**结论** 在本文所选取的样本数据中，灰色关联分析法为分析阳虚质影响因子的适宜技术。本文为中医体质影响因子的研究分析提供了新方法和新思路参考。

【关键词】阳虚质；灰色关联；影响因子；随机森林；模型