

# Using Data Mining to Investigate Correlation between Traditional Chinese Medicine Body Constitution and Postnatal Problems

Winnie W. M. Lam

Department of Mathematics and  
Information Technology  
The Education University of Hong  
Kong  
Hong Kong  
email: winnielam@eduhk.hk

Rebecca W. Y. Lee

Traditional Chinese Medicine (Private  
Practice)  
Hong Kong  
email: novembrelee@gmail.com

Regina W. S. Sit

The Jockey Club School of Public  
Health and Primary Care  
The Chinese University of Hong Kong  
Hong Kong  
email: reginasit@cuhk.edu.hk

**Abstract** – Many mothers experience maternal physical and mental health problems a few weeks after childbirth with unknown causes. To investigate this problem, Traditional Chinese Medicine (TCM) theory is applied to define individuals' body constitution. As constitution refers to the relatively consistent characteristics of individual body structure and body function which affect the susceptibility towards different causing factors and the tendency of disease development, it is useful to predict the postnatal problems with body constitutions. The objective of this paper is to investigate the correlation between prenatal conditions (particularly TCM body constitution) and postnatal health problems with data mining approaches. Data was collected from pregnant women aged 21 to 45 through a standardized Constitution in Chinese Medicine Questionnaire (CCMQ) before childbirth and a face-to-face assessment that was conducted by the TCM practitioner after childbirth. The collected data was analyzed by Pearson's Chi-square test and three benchmark data mining models to discover significant correlation. This study confirms a strong correlation between imbalanced TCM body constitution and postnatal problems. The discovered correlation can help to provide personalized TCM treatment to prevent the potential postnatal problems through an early TCM intervention during pregnancy to regulate the imbalance condition of pregnant women.

**Keywords**—TCM body constitution; postnatal problems; data mining; chi-square test; correlation.

## I. INTRODUCTION

Almost every mother experiences physical and psychological sicknesses after childbirth. Even though the results of prenatal check-up and tests, including ultrasound exams, blood tests, glucose challenge screening, etc., are normal, many women are still suffering from postnatal problems such as prolonged lochia, depression, tiredness, sleep disturbance, headache, constipation, etc. [2]-[4], while the reasons are unknown [6][7]. To ensure both mothers and babies are provided with the best care and protection, in addition to the Western medical check-up and tests, many mothers in Asian countries [8] adopt Traditional Chinese Medicine (TCM) as complementary maternal healthcare during their prenatal period, or even before becoming pregnant. They believe that TCM treatment before childbirth could help them to prevent miscarriage and sicknesses.

TCM is widely used in maternal healthcare in both Asian [8] and Western countries [9] nowadays. TCM practitioners perform predictive, preventive and personalized diagnosis [10][11] to individuals. Treatments could be different for individuals with the same symptom, but the different diagnosis of TCM syndrome differentiation is due to their different body constitution types. TCM body constitution [12] could be divided into nine types: gentleness, qi-

deficiency, yang-deficiency, yin-deficiency, phlegm-damp, damp-heat, blood-stagnation, qi-stagnation and special diathesis, according to Q. Wang [18].

In the theory of body constitution in TCM, every individual has his/her body constitution. Balanced body constitution (i.e., gentleness type) is considered as the most desirable constitution which represents good health [22]. A person with balanced status is the least susceptible to illness [13] because the body can maintain homeostasis physically and psychologically, and people with imbalanced status are considered as unhealthy and prone to sickness or diseases [23] [31]. The imbalanced health status can be caused by the deficiency of "qi" (vital energy), "yang" (masculine or positive principle that is characterized by light/warmth/dryness/activity) or "yin" (feminine or negative principle that characterized by dark/wetness/cold/passivity/disintegration), the build-up of dampness, phlegm or heat, stagnation of "qi" and blood, or special diathesis (sensitive to external stimulants such as drugs, food, smells or seasonal triggers) [24].

Prevention of disease is part of maintaining good health. The aim of this study is to discover the significant correlation between prenatal conditions (especially TCM body constitution) and postnatal problems from the personal profile for each pregnant woman from prenatal to postnatal period, so that postnatal problems can be predicted at an earlier stage. By matching the discovered correlation patterns with the prenatal conditions of pregnant women, it would be possible to predict if a mother will develop particular postpartum diseases or symptoms. This will allow TCM practitioners to provide an intervention procedure to adjust the imbalanced condition to prevent certain postnatal problems.

The rest of the paper is organized as follows. Section II describes the methods of the data collection, preparation and modelling. Section III presents the experimental results. Finally, Sections IV and V are the discussion and conclusion.

## II. METHODS

### A. Selection of Participants

The recruitment of participants was started in May, 2013 in community TCM clinics in Hong Kong. The selection criteria were women who are over 18 years old and living in Hong Kong for more than 10 years. To avoid outliers or external factors, women with serious health problems including HIV/AIDS, cancer, viral infections (e.g., Ebola, Zika virus, etc.) within a year were excluded.

### B. Building Health Profiles

Data was collected from 1) prenatal period (i.e., the time between conception and childbirth) and 2) postnatal period

(i.e., six weeks after childbirth) to build the health profiles that include the maternal health conditions of participants.

### 1) Prenatal Profile

In general medical practice, the TCM practitioner can determine the body constitution of patients by four diagnostic methods including observation (face, body form, tongue, etc.), smelling and hearing (odours from mouth/body, loudness of speaking voice), inquiry (symptoms) and palpitation (pulse) [17]. To classify each participant into one or multiple body constitutions systematically, each patient needs to complete a standardized Constitution in Chinese Medicine Questionnaire (CCMQ) that was developed by Q. Wang [18] and China Association of Chinese Medicine [19] with the diagnosis and validation of the TCM practitioner. The CCMQ is a self-rating scale questionnaire that includes 60 questions related to a variety of prenatal conditions (e.g., cold/warm hands and feet, anxiety level, dark circles around the eyes, dry mouth, etc.). The participants should complete the questionnaire in *Google Forms* according to their physical and mental health condition in the prenatal period.

#### a) CCMQ Score

A set of CCMQ scores  $S = \{s_1, s_2, s_3, \dots, s_9\}$  that are corresponding to the nine types of TCM body constitution ( $c_1$ : gentleness,  $c_2$ : qi-deficiency,  $c_3$ : yang-deficiency,  $c_4$ : yin-deficiency,  $c_5$ : phlegm-damp,  $c_6$ : damp-heat,  $c_7$ : blood-stagnation,  $c_8$ : qi-stagnation and  $c_9$ : special diathesis) respectively are calculated from the CCMQ by following the equations in [20]. After calculating the scores, all participants are classified into one or multiple TCM body constitution types based on Table I.

TABLE I. CLASSIFICATION OF TCM BODY CONSTITUTION TYPES

TCM body constitution	Conditions <sup>a</sup>	Results <sup>a</sup>
<b>Balanced</b> ( $c_1$ : gentleness)	$s_1 \geq 60$ and $s_n < 30$	$c_1$
	$s_1 \geq 60$ and $s_n < 40$	Marginal $c_1$
	$s_n \geq 40$	Not $c_1$
<b>Imbalanced</b> ( $c_2$ : qi-deficiency $c_3$ : yang-deficiency $c_4$ : yin-deficiency $c_5$ : phlegm-damp $c_6$ : damp-heat $c_7$ : blood-stagnation $c_8$ : qi-stagnation $c_9$ : special diathesis)	$s_n \geq 40$	$c_n$
	$30 \leq s_n \leq 39$	Tends to $c_n$
	$s_n < 30$	Not $c_n$

<sup>a</sup>. where  $n = 2$  to  $9$

#### b) Classification of TCM Body Constitution

In the classification of TCM body constitution types, we can identify the likelihood (in terms of CCMQ score) of each body constitution type of each participant according to Table I. Each participant may have one or more dominant body constitution types, or even have a conflict between balanced and imbalanced body constitutions. Here are three sample cases:

Case 1: Given a participant  $P_1$  with CCMQ scores  $S_1 = \{75, 56, 27, 25, 12, 15, 20, 18, 10\}$ .  $s_1$  is higher than 60, but  $s_2$  is also higher than 40. Hence,  $P_1$  is classified as  $c_2$  instead of  $c_1$ .

Case 2: Given a participant  $P_2$  with CCMQ scores  $S_2 = \{75, 16, 27, 25, 32, 25, 10, 18, 10\}$ . As  $s_1$  is higher than 60,  $s_5$  is higher than 30, and  $s_n$  is less than 40 where  $n = 2$  to  $9$ ,  $P_2$  is classified as  $c_1$  marginally. Also,  $s_5$  is between 30 and 39, so

$P_2$  has a mixture of two body constitution types: marginal  $c_1$  and tends to  $c_5$ .

Case 3: Given a participant  $P_3$  with CCMQ scores  $S_3 = \{59, 16, 27, 25, 12, 15, 20, 18, 10\}$ .  $s_1$  is less than 60, and  $s_n$  is less than 30 where  $n = 2$  to  $9$ ,  $P_3$  is classified as undefined body constitution.

### 2) Postnatal Profile

All participants who completed the CCMQ in the first stage should arrange a face-to-face assessment with a registered TCM practitioner after childbirth. The TCM practitioner needs to complete the postnatal profile that includes two parts: a) demographics and maternal health information, and b) physical and mental postnatal health problems within six weeks after childbirth.

#### a) Demographics and Maternal Health Information

In the first part, the basic demographic data, such as age, occupation and blood type, are collected, and the other questions are related to the menses cycle and volume, ovulation, delivery method, and the number of fetuses of the participants.

#### b) Physical and Mental Postnatal Health Problems

In the second part, postnatal health problems including prolonged lochia (i.e., vaginal discharge after giving birth containing blood, mucus, and uterine tissue over 21 days), tiredness, abnormal sweating, etc. are examined. Since all the questions are TCM-specific, some factors, such as tongue appearance and pulse patterns, should be examined face-to-face by a TCM practitioner.

### C. Data Preparation

Firstly, the data of the prenatal and postnatal profiles was merged and stored into a database, which includes 132 rows (i.e., records) and 210 columns (i.e., attributes). Secondly, incorrect and inconsistent data was identified by the TCM practitioner, and the identified fields and records were either corrected or removed from the data set. Lastly, the continuous data, such as age, age of first menses, period cycle and week of childbirth, were discretized into intervals, as advised by the TCM practitioner.

### D. Identify Dependent and Independent Attributes

Dependent and independent attributes [21] were identified from our collected data, as shown in Table II. The dependent attributes represent the outcome that we want to predict (i.e., postnatal problems), and the independent attributes represent the input of potential causes (e.g., demographics, prenatal conditions, TCM body constitution, etc.).

### E. Data Modeling

The correlation between prenatal health conditions (especially TCM body constitution) and postnatal problems can be discovered by supervised data mining models. We used an open-source data mining tool named *Orange*, which includes these three supervised data mining models: decision tree [26], Support Vector Machine (SVM) [27] and Artificial Neural Networks (ANN) [28][29] for discovering significant patterns from the collected maternal data and comparing the effectiveness of different data mining models on TCM data analysis. The *Orange* version 3.14 was installed. To obtain the best accuracy and results, the parameters of the decision tree, SVM and ANN were set as described below.

In the decision tree modelling, the "Tree" widget was used. The parameters of the minimum number of instances in

leaves, the smallest number of instance of splitting subsets and maximal tree depth were set to 2, 5 and 100, respectively. In the SVM modelling, Radial basis function (RBF) kernel was used, and the parameters of C and  $\epsilon$  were set to 1.00 and 0.1, respectively to obtain the best accuracy. In the ANN modelling, the number of neurons per hidden layer was set to 100, and the parameters of activation, solver and alpha were set to ReLu (rectified linear unit function), Adam (stochastic gradient-based optimizer) and 0.0001 respectively.

TABLE II. DEPENDENT AND INDEPENDENT ATTRIBUTES

Dependent attributes (postnatal problems)	Independent attributes	Remarks	
1. Tiredness	1. Age		
2. Unhealthy face colour	2. Occupation		
3. Prolonged lochia	3. Living environment		
4. Excessive sweating	4. Blood type		
5. Dry mouth	5. Genetic/ congenital disease		
6. Feeling annoyed	6. Chronic disease		
7. Joint pain	7. Gynecological disease		
8. Feeling of anger	8. Age of first menses	Related to menses	
9. Feeling hot/ Hot flash	9. Regular period		
10. Emotional depression	10. Days of period cycle		
11. Fear of cold/ wind	11. Days of menstruation		
12. Excessive dreaming	12. Menses volume		
13. Body pain	13. Menses colour		
14. Shortness of breath	14. Menses with blood clot		
15. Dizziness	15. Dysmenorrhea		
16. Headache	16. Number of pregnancy		Related to pregnancy and delivery
17. Flat feeling in mouth	17. Number of abortion		
18. Bright yellow urine	18. Number of miscarriage		
19. General fatigue	19. Number of fetuses		
20. Tinnitus	20. Expected date of delivery (EDD)		
21. Dry/ hard stool	21. Date of childbirth		
22. Insatiable hunger	22. Week of childbirth		
23. Constipation	23. Way of delivery		
24. Excessive urine	24. Curettage		
25. Excessive belching	25. Medication	Taken during pregnancy	
26. Loose stool	26. Supplement		
27. Poor appetite	27. Number of days in hospital	After childbirth	
28. Fever	28. Follow traditional rituals		
29. Palpitation	29. Breastfeeding	Obtained from CCMQ	
30. Insomnia	30. TCM body constitution types		
31. Bitter taste in mouth			
32. Epigastric distension			
33. Rectal tenesmus			
34. Stomachache			
35. Gastric excretion (from mouth)			
36. Frequent nighttime urination			

To prevent over fitting, the validation was carried out with 5-fold cross validation due to the small sample size. The data was split into a training set (80%) for building the model and test set (20%) for validating the built model. The data mining steps are summarized here:

- Step 1: Import the collected data with the “File” widget.
- Step 2: Define the input and output attributes by the “Select Columns” widget.
- Step 3: Feed the selected data to different data mining models: “Tree” (i.e., decision tree), “SVM” and “Neural Network” (i.e., ANN).
- Step 4: Perform cross-validation accuracy estimation with the “Test & Score” widget.
- Step 5: Predict the unknown output with the “Predictions” widget.
- Step 6: Repeat Step 2 to remove the outliers and adjust the parameter settings to obtain better accuracy.

### III. RESULTS

#### A. Distributions

##### 1) Demographics

A total of 132 pregnant women aged 21 to 45 were recruited from May, 2013 to May, 2017. Participants aged 31 to 35 was the majority group (55.30%), followed by the age groups 26 to 30 (16.67%) and 36 to 40 (18.18%). Around half of the participants (56.8%) were working indoors, and some of them were not working (16.67%).

##### 2) Maternal Health Information

According to the pregnancy information, 59.09% of the participants were at their first pregnancy, and over 90% of them had one fetus. Although most of the mothers were reported as healthy and had regular menses cycle and normal menses volume, there were 69.70% of them who mentioned that they were suffering from different kinds of discomforts during pregnancy. To reduce the risk of miscarriage or discomfort during pregnancy, most participants (88.64%) took Western supplements such as folic acid, Materna, and a few of them (37.12%) also took Chinese medicine. When they were sick, half of them did not take either Western or Chinese medicine because they were afraid the medicine would influence the health of their fetuses.

##### 3) Delivery Information

The delivery dates of most of our participants (39.39%) were in winter, while the others who delivered in summer were the least (13.64%). Over 90% of them had full-term pregnancy (i.e., delivered within 37 to 42 weeks), and natural childbirth was the main type (61.36%) of delivery. Most of them (78.79%) were able to leave the hospital after childbirth within 4 days, and only a few (4.55%) were staying in the hospital for more than 6 days. After the delivery, 72.73% of the participants followed the traditional rituals of TCM (e.g., no hair wash, use TCM or ginger for body wash, avoid all cold and raw food and drinks, etc.) during their postnatal period.

##### 4) Physical and Mental Postnatal Health Problems

Within the postnatal period, our participants were suffering from different postnatal health problems. We found that 92.42% of them suffered from four or more different health problems after childbirth, and the occurrence of 16 common postnatal TCM symptoms are shown in Table III (in descending order).

##### 5) TCM Body Constitution

In the prenatal profile, participants were classified into nine types of TCM body constitution: gentleness, qi-deficiency, yang-deficiency, yin-deficiency, phlegm-damp, damp-heat, blood-stagnation, qi-stagnation and special diathesis according to the equations in Table I. Except for the gentleness type that is considered as balanced body constitution, the other types are considered as imbalanced. In

our data, 100 women were classified as imbalanced body constitution, 28 were balanced, and 4 were unclassified.

By considering the participants who were belonging to a certain body constitution type obviously (i.e.,  $c_n$  is “Yes”), the top three TCM body constitution types are Yin-deficiency (37.12%), Yang-deficiency (35.61%) and blood-stagnation (35.61%).

It is important to note that not all individuals would have a single body constitution type; it is common to have a mixed body constitution (i.e., two or more body constitution types). Among 104 participants who were considered as having imbalanced body constitution, 30.77% participants had single body constitution type, 65.38% of them had two or more, and only a small proportion (3.85%) were unclassified.

TABLE III. COMMON POSTNATAL TCM SYMPTOMS

Postnatal symptoms	%	Postnatal symptoms	%
Tiredness	80.30	Feeling hot/ Hot flash	44.70
Unhealthy face colour	68.18	Emotional depression	44.70
Prolonged lochia <sup>a</sup>	66.67	Fear of cold/ wind	43.94
Excessive sweating	60.61	Excessive dreaming	43.94
Dry mouth	59.09	Body pain	34.85
Feeling annoyed	59.09	Shortness of breath	31.82
Joint pain	53.79	Dizziness	31.82
Feeling of anger	48.48	Headache	31.06

<sup>a</sup>. The number of days of lochia clearance is more than 21 days

B. Correlation between TCM Body Constitution and Postnatal Problems

1) Statistical Test for Independence

Pearson’s Chi-square test [30] was used to measure the correlations between TCM body constitution types and postnatal TCM symptoms, and the null hypothesis was applied to prove whether the occurrence of these two attributes was statistically independent. After carrying out the test in a statistical software package SPSS, we discovered that some postnatal symptoms from Table IV did not occur by chance, but were statistically dependent on certain TCM body constitution types. The significant pairs of attributes with a 95% confidence interval (i.e., p-value < 0.05) were listed in Table IV with the results of Cramer’s V that indicates the strength statistic. The coefficient of Cramer’s V ranges from 0 (no association) to 1 (strongest association). With the degree of freedom  $df = 2$ , the strength is considered as small, medium and large for the values 0.07, 0.21 and 0.35, respectively [32].

Within the 45 significant pairs of attributes, 16 pairs were statistically significant (i.e., significance level < 0.01) with medium/ strong strength in Cramer’s V. However, the Chi-square test with Cramer’s V can only determine the statistical significance between two attributes, but how exactly they are related to each other is unknown (e.g., Is qi-deficiency/ tends to qi-deficiency/ non-qi-deficiency related to emotional depression or not?). Thus, we need to use decision tree modelling to discover detailed relationships.

2) Decision Tree Modeling

From the results of Pearson’s Chi-square test in Table IV, we observed that some TCM body constitution types are highly correlated to certain postnatal symptoms. To further investigate the detailed relationships between these two attributes, we selected a significant pair of attributes (qi-deficiency and dizziness) with 99% confidence interval (i.e., p-value is less than 0.01) to demonstrate how the decision tree is used to discover the correlations in the form of tree and rules.

First, the collected data was imported into Orange. Next, it was connected to the other widgets (i.e., Select Columns, Tree, Test & Score and Tree Viewer). In the “Select Columns” widget, the TCM body constitution “qi-deficiency” and the postnatal symptom “dizziness” were set as input and output, respectively. Finally, we obtained the decision tree model (in Figure 1) with an overall classification accuracy of 72.7%.

TABLE IV. SIGNIFICANT RESULTS OF PEARSON’S CHI-SQUARE TEST

TCM body constitution types	Attributes	Significance level (< 0.01)	Cramer’s V Result (> 0.35)
	Postnatal symptoms		
qi-deficiency	Excessive sweating	0.045	
	Tiredness	0.026	
	Shortness of breath	0.013	
	Feeling hot/ Hot flash	0.036	
	Headache	<b>0.004</b>	0.292
	Dizziness	<b>0.000065</b>	<b>0.382</b>
	Emotional depression	<b>0.002</b>	0.309
	Excessive urine	0.011	
	General fatigue	0.04	
	Joint pain	0.012	
Yang-deficiency	Body pain	<b>0.009</b>	0.267
	Dizziness	0.041	
	Fear of cold/wind	<b>0.000245</b>	<b>0.355</b>
Yin-deficiency	Stomachache	0.044	
	Bitter taste in mouth	0.031	
phlegm-damp	Dry mouth	<b>0.002</b>	0.301
	Dry/hard stool	0.035	
	Dizziness	0.013	
damp-heat	Bitter taste in mouth	<b>0.000011</b>	0.345
	Flat feeling in mouth	0.012	
	Feeling annoyed	0.03	
	Bright yellow urine	0.043	
	Frequent nighttime urination	<b>0.00004</b>	<b>0.392</b>
blood-stagnation	Dizziness	0.024	
	Bitter taste in mouth	<b>0.00004</b>	0.345
	Emotional depression	0.042	
	Frequent nighttime urination	0.01	
	Shortness of breath	0.035	
	Dizziness	<b>0.003</b>	0.299
qi-stagnation	Bitter taste in mouth	0.033	
	Dry mouth	0.027	
	Flat feeling in mouth	<b>0.006</b>	0.276
	Feeling annoyed	<b>0.001</b>	0.317
	Feeling of anger	0.013	
	Face colour	0.049	
	Fever	<b>0.001</b>	0.325
special diathesis	Dizziness	<b>0.001</b>	0.337
	Excessive belching	0.014	
	Bright yellow urine	0.025	
	Body pain	0.04	
	Dizziness	<b>0.005</b>	0.283
	Bitter taste in mouth	0.012	
	Feeling annoyed	0.033	
	Rectal tenesmus	0.024	
Prolonged lochia	<b>0.003</b>	0.295	

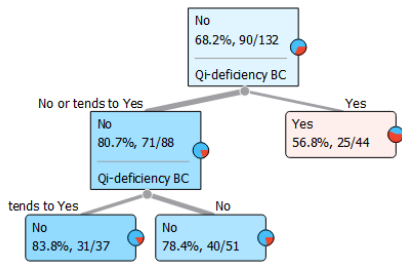


Figure 1. Relationships between qi-deficiency and dizziness in the decision tree

In this output of the decision tree model, we discovered three relationships between qi-deficiency and dizziness by following the paths from the root node to the leaf nodes. These paths are represented in the form of IF-THEN rules, where  $n$  refers to the number of participants who matched the following rules:

**Rule 1:** IF qi-deficiency = “tends to Yes” THEN dizziness = “No” ( $n = 31$ , accuracy = 83.8%)

**Rule 2:** IF qi-deficiency = “No” THEN dizziness = “No” ( $n = 40$ , accuracy = 78.4%)

**Rule 3:** IF qi-deficiency = “Yes” THEN dizziness = “Yes” ( $n = 25$ , accuracy = 56.8%)

In addition to the one-to-one correlation between qi-deficiency and dizziness discovered by Pearson’s Chi-square test, a decision tree can define multiple inputs to find whether the output attribute (i.e., dizziness) is correlated to one or more input attributes. From Table IV, we found that dizziness is not only strongly correlated to qi-deficiency, but it is also correlated to the other six imbalanced TCM body constitution types. Instead of setting one input attribute in the decision tree, we set all eight imbalanced TCM body constitution types as input to predict the postnatal dizziness.

Among the eight imbalanced TCM body constitution types, qi-deficiency was selected as the root node (i.e., best predictor) with the highest information gain. The leaf node with the highest classification accuracy and number of samples was extracted from the rightmost part of the tree and shown in Figure 2. The leaf node shows that all women with multiple TCM body constitution types: qi-deficiency, special diathesis and tends to phlegm-damp were suffering from dizziness in the postnatal period. This proved that dizziness is related to multiple TCM body constitution types instead of one only. Since there were over 60% of women with multiple body constitution types and different prenatal conditions, a decision tree with multiple input attributes is the better way to predict postnatal problems.

In order to discover more possible prenatal factors that cause the postnatal problems, a decision tree and two other supervised data mining algorithms were applied to our data, including the attributes of TCM body constitution, demographics, maternal health information and delivery information, collected from the prenatal period to predict the postnatal problems. The independent (i.e., input) and dependent (i.e., output) attributes are listed in Table II.

**C. Significant Patterns with Data Mining Algorithms**

Supervised data mining was used to discover significant patterns from prenatal and postnatal data. The classification accuracies of predicting 36 postnatal problems, as shown in Table II, are reported in Figure 3. The classification accuracy of each postnatal problem (target attribute) in the decision tree, SVM and ANN are indicated by blue triangle dot, red circle dot and grey cross, respectively.

Half of the postnatal problems can be predicted with classification accuracy over 70%, and the postnatal problem no. 36 (i.e., frequent nighttime urination) has the highest classification accuracy of 92.27%. Overall, the average accuracies of the three supervised data mining algorithms are above 64%, and SVM has the highest average accuracy of 73.07%.

To predict the postnatal problem of frequent nighttime urination, the ROC curves of the decision tree, SVM and neural network are given in Figure 4. It shows the performance of neural network is the best among three with higher AUC, followed by decision tree, and lastly SVM.

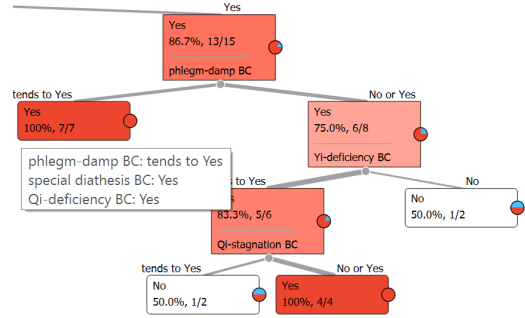


Figure 2. Leaf node with the highest accuracy rate to predict dizziness

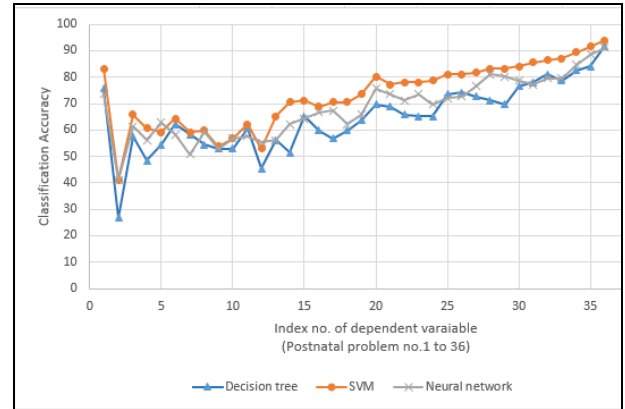


Figure 3. Classification accuracies of 36 postnatal problems

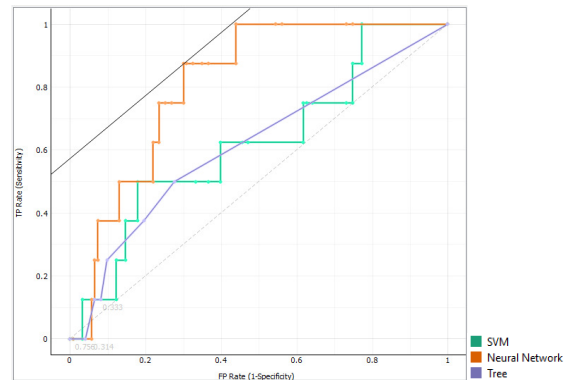


Figure 4. ROC curves of decision tree, SVM and neural network

After consulting the TCM practitioner, four postnatal problems, including dizziness, prolonged lochia, tiredness, and emotional depression, were identified as important to clinical care after pregnancy. We used a decision tree to discover the correlations between these postnatal problems and prenatal factors because it can show the results in terms of readable rules which are understandable by TCM practitioners.

### 1) Classification Rules of Dizziness

As mentioned earlier, dizziness is highly correlated to multiple TCM body constitution types. From *Rule 1.1* to *Rule 1.4*, we can predict if a woman will suffer from dizziness after pregnancy related to five prenatal factors: 1) the number of TCM body constitution type, 2) TCM body constitution types (including qi-stagnation, Yang-deficiency, Yi-deficiency and special diathesis), 3) blood type, 4) living environment, and 5) number of days of menstruation before pregnancy.

*Rule 1.1:* IF the number of TCM body constitution types > 5 THEN postnatal problem = “dizziness” ( $n = 13$ , accuracy = 86.7%)

*Rule 1.2:* IF the number of TCM body constitution types  $\leq 5$  AND blood type = “AB, O or unknown” AND living environment = “Above the hill” THEN postnatal problem = “dizziness” ( $n = 5$ , accuracy = 100%)

*Rule 1.3:* IF the number of TCM body constitution types  $\leq 5$  AND days of menstruation  $\leq 7$  AND living environment  $\neq$  “above the hill”, THEN postnatal problem  $\neq$  “dizziness” ( $n = 85$ , accuracy = 80.2%)

*Rule 1.4:* IF the number of TCM body constitution types  $\leq 5$  AND TCM body constitution type = “qi-stagnation, Yang-deficiency, Yi-deficiency or special diathesis” AND days of menstruation  $\leq 7$  AND living environment  $\neq$  “above the hill”, THEN postnatal problem  $\neq$  “dizziness” ( $n = 18$ , accuracy = 100%)

By comparing *Rule 1.3* and *Rule 1.4*, we observed that when the additional factor, TCM body constitution, is considered when predicting dizziness, it can yield higher accuracy. Hence, the TCM body constitution is an important factor that can affect the accuracy of predicting dizziness.

### 2) Classification Rules of Prolonged Lochia

Prolonged lochia is the third most common postnatal problem in this study, and the abnormal duration of bleeding can be suspected as infection. According to *Rule 2.1* to *Rule 2.4*, we can detect this problem in the prenatal period by observing seven prenatal factors: 1) TCM body constitution type (special diathesis and qi-stagnation), 2) days of period cycle, 3) traditional rituals (avoid all cold/raw food), 4) tongue coating, 5) EDD, 6) blood type and 7) dysmenorrhea before or after menstruation.

*Rule 2.1:* IF TCM body constitution type (special diathesis)  $\neq$  “tends to Yes” AND avoid all cold/raw food = “No” THEN postnatal problem = “prolonged lochia” ( $n = 53$ , accuracy = 81.5%)

*Rule 2.2:* IF TCM body constitution types  $\neq$  “qi-stagnation or special diathesis” AND avoid all cold/raw food = “No” AND days of period cycle > 31 THEN postnatal problem = “prolonged lochia” ( $n = 25$ , accuracy = 100%)

*Rule 2.3:* IF TCM body constitution type (special diathesis) = “tends to yes” AND EDD = “spring or summer” THEN postnatal problem  $\neq$  “prolonged lochia” ( $n = 7$ , accuracy = 100%)

*Rule 2.4:* IF TCM body constitution (special diathesis)  $\neq$  “tends to yes” AND avoid all cold/raw food = “Yes” AND blood type = “O or unknown” AND dysmenorrhea = “occurred before or after menstruation” THEN postnatal problem  $\neq$  “prolonged lochia” ( $n = 7$ , accuracy = 100%)

### 3) Classification Rules of Emotional Depression

Emotional depression is one of the top 10 postnatal problems, and it is a type of mood disorder associated with childbirth. It can negatively affect the mother, her family and

newborn child. This problem can be detected by monitoring five factors: 1) TCM body constitution type, 2) blood vessel under the tongue, 3) days of menstruation, 4) menses, and 5) dysmenorrhea based on *Rule 4.1* to *Rule 4.4*:

*Rule 4.1:* IF days of menstruation > 4 AND age of the first menses  $\neq$  “11–12” AND TCM body constitution type (qi-deficiency) = “yes or tends to yes” THEN postnatal problem = “Emotional depression” ( $n = 31$ , accuracy = 81.6%)

*Rule 4.2:* IF menses with blood clot = “no” AND menses colour  $\neq$  “bright red or pale red” AND age of first menses = “11–12” AND TCM body constitution type (qi-deficiency) = “yes or tends to yes” THEN postnatal problem = “Emotional depression” ( $n = 7$ , accuracy = 87.5%)

*Rule 4.3:* IF TCM body constitution type = “complex, unclassified, Yang-deficiency or Yi-deficiency” AND qi-deficiency = “no” THEN postnatal problem  $\neq$  “Emotional depression” ( $n = 29$ , accuracy = 90.6%)

*Rule 4.4:* IF Menses colour = “bright red or pale red” AND Age of first menses = “11–12” AND TCM body constitution type (qi-deficiency) = “yes or tends to yes” THEN postnatal problem  $\neq$  “Emotional depression” ( $n = 5$ , accuracy = 100%)

## IV. DISCUSSION

In this study, we found that 92.42% of the women suffered from four or more different health problems after childbirth. Women with “gentleness” body constitution type tended to have less postnatal discomfort; the finding was consistent with the TCM theory that individuals with balanced body constitution are more likely to stay healthy even in the stressful postnatal period. Contrarily, women with imbalanced TCM body constitution were found to have more postnatal problems, among which qi-deficiency and blood stagnation had the worse postnatal profiles.

In our analysis, several factors were found to predict the outcome of dizziness. However, some factors, such as the week of childbirth, duration of period cycle and menses volume, were unmodifiable. Other factors, namely traditional rituals and the intake of calcium supplements could be changed easily through prenatal education and advice. Additionally, although TCM body constitution is constant and not easily changed at once, it can be adjusted in the long run through lifestyle modification, nutrition, exercise and herbal medications. Therefore, we believe that by knowing individuals’ body constitution and their prenatal habits could potentially improve women’s postnatal health.

There are several limitations to our study. First, the study was conducted in Hong Kong, and only women with Chinese ethnicity were invited; therefore, the findings may not be generalizable. Second, the correlations between TCM body constitution and postnatal problems were not one-to-one relationships; in both Pearson’s Chi-square test and the three supervised data mining algorithms (SVM, decision tree and artificial neural networks), we can only set one attribute (one of the postnatal problems) as output at one time. Third, TCM body constitution is potentially adjustable before pregnancy; a future study can be conducted on the prediction of the other disease entities [13][14], especially emotional depression. Additional factors can be collected from Western medical records for building a better predictive model.

## V. CONCLUSION

In this study, we discovered that women with imbalanced TCM body constitution were more likely to suffer from postnatal health problems. Some body constitution types and prenatal habits were predictive of certain common and important postnatal symptoms, which were believed to be modifiable. Supervised data mining has been demonstrated as a useful way to discover correlations between multiple prenatal conditions (especially body constitutions) and postnatal problems. After discovering the significant correlations in terms of classification rules, the future step is to predict postnatal problems before childbirth. By using the associations between body constitution and postnatal problems in Table IV and the classification rules in Section III C), the TCM practitioners can apply early interventions, such as medicine, acupuncture, suggesting nutrition or supplements, to adjust individuals' body constitution and prevent the postnatal problems during the prenatal period, or even before pregnancy.

## REFERENCES

- [1] D. Neagos, R. Cretu, R. C. Sfetea and L. C. Bohiltea, "The Importance of Screening and Prenatal Diagnosis in the Identification of the Numerical Chromosomal Abnormalities," *Mædica*, 6(3), pp. 179-184, 2011.
- [2] S. Brown and J. Lumley, "Maternal health after childbirth: results of an Australian population based survey," *An International Journal of Obstetrics and Gynaecology*, vol. 105 (2), pp. 156-161, February 1998.
- [3] J. Y. Lee and J. Y. Hwang, "A study on postpartum symptoms and their related factors in Korea," *Taiwanese Journal of Obstetrics and Gynecology*, vol. 54 (4), pp. 355-363, August 2015.
- [4] S. Aksu, F. G. Varol and N. H. Sahin, "Long-term postpartum health problems in Turkish women: prevalence and associations with self-rated health," *Contemporary Nurse*, 53:2, pp. 167-181, 2017.
- [5] A. Rudman and U. Waldenström, "Critical views on postpartum care expressed by new mothers," *BMC Health Services Research*, 7, pp. 178, 2007.
- [6] N. T. Hatfield, "Introductory Maternity and Pediatric Nursing," Lippincott Williams & Wilkins, pp. 405, Nov 2013.
- [7] L. I. Alasoom and M. R. Koura, "Predictors of Postpartum Depression in the Eastern Province Capital of Saudi Arabia," *Journal of Family Medicine and Primary Care*, vol. 3(2), pp. 146-150, 2014.
- [8] Q. Y. Jiang, J. Li, L. Zheng, G. H. Wang, and J. Wang, "Constitution of traditional chinese medicine and related factors in women of childbearing age," *Journal of the Chinese Medical Association*, 81(4), pp. 358-365, 2018.
- [9] S. Lukman, Y. He, and S. C. Hui, "Computational methods for traditional Chinese medicine: a survey," *Computer methods and programs in biomedicine*, 88(3), pp. 283-294, 2007.
- [10] W. Wang, A. Russell, Y. Yan, "Traditional Chinese medicine and new concepts of predictive, preventive and personalized medicine in diagnosis and treatment of suboptimal health," *EPMA Journal*, 5(1), pp. 12, 2014.
- [11] Y. Wang, Y. Dai, F. Guo, S. Li, "Sensitivity-based data selection for predicting individual's sub-health on TCM doctors' diagnosis," *IEEE International Conference on Systems, Man, and Cybernetics, Anchorage, AK*, pp. 896-901, 2011.
- [12] Y. B. Zhu, Q. Wang, H. Origasa, "Evaluation on reliability and validity of the constitution in chinese medicine questionnaire (CCMQ)," *Chinese Journal of Behavioral Medical Science*, 16(7), pp. 651-654, 2007.
- [13] H. You, T. Zhang, W. Feng, and Y. Gai, "Association of TCM body constitution with insulin resistance and risk of diabetes in impaired glucose regulation patients," *BMC complementary and alternative medicine*, 17(1), pp. 459, 2017.
- [14] Y. Zhu, et al., "Association between nine types of TCM constitution and five chronic diseases: a correspondence analysis based on a sample of 2,660 participants," *Evidence-Based Complementary and Alternative Medicine*, 9439682, 2017.
- [15] C. I. Tsai, Y. C. Su, S. Y. Lin, I. T. Lee, C. H. Lee & T. C. Li, "Reduced health-related quality of life in body constitutions of Yin-Xu, and Yang-Xu, stasis in patients with type 2 diabetes: taichung diabetic body constitution study," *Evidence-Based Complementary and Alternative Medicine*, 309403, 2014.
- [16] J. Iavindrasana, G. Cohen, A. Depeursinge, H. Müller, R. Meyer & A. Geissbuhler, "Clinical Data Mining: a Review," *Methods of Information in Medicine*, pp. 121-133, 2009.
- [17] L. Zhang, "A TCM Way to be Healthy, Inside and Out," FriesenPress, 2017.
- [18] Q. Wang, "Classification and diagnosis basis of nine basic constitutions in chinese medicine," *Journal of Beijing University of Traditional Chinese Medicine*, 28(4), pp. 1-8, 2005.
- [19] M. Li, S. Mo, Y. Lv, Z. Tang and J. Dong, "A Study of Traditional Chinese Medicine Body Constitution Associated with Overweight, Obesity, and Underweight," *Evidence-Based Complementary and Alternative Medicine*, 7361896, 2017.
- [20] W. Wong, C. L. K. Lam, V. T. Wong, Z. M. Yang, E. T. C. Ziea & A. K. L. Kwan, "Validation of the Constitution in Chinese Medicine Questionnaire: Does the Traditional Chinese Medicine Concept of Body Constitution Exist?," *Evidence-Based Complementary and Alternative Medicine*, 481491, 2013.
- [21] I. Kononenko, "Estimating attributes: Analysis and extensions of RELIEF," *European Conference on Machine Learning*, pp. 171-182, 1994
- [22] A. Al-Shura, "Integrative Cardiovascular Chinese Medicine: A Prevention and Personalized Medicine Perspective," Academic Press, pp. 39-56, 2014
- [23] R. Y. Chan, and W. T. Chien, "Concepts of body constitution, health and sub-health from traditional Chinese medicine perspective," *World J Transl Med*, 2(3), pp. 56-66, 2013.
- [24] G. Maciociu, "The Practice of Chinese Medicine," New York, Churchill Livingstone, 1994.
- [25] J. Han, J. Pei and M. Kamber, "Data mining: concepts and techniques," Elsevier, 2011.
- [26] S. R. Safavian and D. Landgrebe, "A survey of decision tree classifier methodology," *IEEE Transactions on System, Man, and Cybernetics*, 21(3), pp. 660-674, 1991.
- [27] H. Bhavsar and M. H. Panchal, "A review on support vector machine for data classification," *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 1(10), pp. 185, 2012.
- [28] F. Rossi and N. Villa, "Support vector machine for functional data classification," *Neurocomputing*, 69(7-9), pp. 730-742, 2006.
- [29] R. J. Schalkoff, "Artificial Neural Networks," McGraw-Hill, 1997.
- [30] H. Scheffé, "The Relation of Control Charts to Analysis of Variance and Chi-Square Tests," *Journal of the American Statistical Association*, 42(239) pp. 425-431, 1947.
- [31] R.Y. Chan and W.T. Chien, "Concepts of body constitution, health and sub-health from traditional Chinese medicine perspective," *World J Transl Med*, 2, pp. 56-66, 2013.
- [32] J. Cohen, "Statistical power and analysis for the behavioral sciences," 2nd ed. Hillsdale, NJ: Lawrence Erlbaum Associates, pp. 79-80, 1988.