World Scientific
www.worldscientific.com

BUILDING AND EVALUATING AN ANIMAL MODEL FOR SYNDROME IN TRADITIONAL CHINESE MEDICINE IN THE CONTEXT OF UNSTABLE ANGINA (MYOCARDIAL ISCHEMIA) BY SUPERVISED DATA MINING APPROACHES

SHUZHEN GUO,*,§ JIANXIN CHEN,*,§ HUIHUI ZHAO,*,§ WEI WANG,*,‡
JIANQIANG YI,† LEI LIU,* QIGE QI,* RENQUAN LIU,*
QI QIU* and YONG WANG*

*Beijing University of Chinese Medicine
11 Bei San Huan Dong Lu, Chao Yan District
Beijing 100029, P. R. China

†Institute of Automation, Chinese Academy of Science
Beijing 100190, P. R. China
‡wangwei@bucm.edu.cn

Building an animal model for a disease is a better avenue to understand the inner mechanism of it. Traditional Chinese Medicine accumulated much practical experience and a large amount of literature to heal diseases during the past 3000 years. However, as there is no available animal model for TCM research because syndrome, the core of TCM theory, it is hard to be diagnosed from animals. In this paper, we present a novel strategy to build and evaluate an animal model for syndrome in TCM in the context of a disease. We first carried out a clinical epidemiology survey for a syndrome (Blood stasis syndrome, BSS) diagnosed by TCM experts in the context of a disease (Unstable angina, UA). Meanwhile, the blood samples of patients included in the survey were collected and measured as physical and chemical specifications by laboratory examinations. Alternatively, we used supervised data mining methods to build association between the specifications and the syndrome in the context of UA. The accuracy of classification was used to evaluate performance of the association built. Finally, we built an animal model for myocardial ischemia and validated the model by established diagnosis criterion of myocardial ischemia. Furthermore, the built association was used to evaluate whether an animal is with BSS. The results indicated that the strategy successfully evaluates and separates the animal model for syndrome in TCM from the counterpart for myocardial ischemia. The novel strategy presented in the paper provides a better insight to understand the nature of syndrome in TCM and pave a basis for personalized therapies of UA.

Keywords: Unstable Angina; Blood Stasis Syndrome; Traditional Chinese Medicine; Data Mining; Animal Model; Myocardial Ischemia; Classification; Radial Basis Function Network.

‡Corresponding author.
§These authors contributed equally to the work.

1. Introduction

Building an animal model for a disease is a better avenue not only to investigate the disease state in ways which would be inaccessible in human patients with the disease, but also to test new drugs for the treatment of the disease. Unstable angina (UA) is a serious disease that causing more than 1 million deaths annually in China. It accounts for heavy burden not only on patients and their families but also on society. Therefore, an increasing number of UA patients pose a major challenge to the entire medical community. The animal model for UA is usually simulated by animal model for myocardial ischemia since the core pathology of UA is myocardial ischemia. There are three widely used methods to reproduce myocardial ischemia animal models: an occlusion of the coronary artery by microembolization; coronary artery ligation; and Ameroid constrictor or coling/gelefoam.[1] Among them, Ameroid constrictor is best used to induce chronic myocardial ischemia as compared with other methods due to its progressive occlusion. Ameroid constrictors have been used in many species of animals. However, porcine hearts have a coronary anatomy most similar to the counterpart of human hearts. They seem to develop much less collateral circulation than dogs after induction of coronary occlusion, which makes them more attractive for the reproduction of chronic myocardial ischemia.[2]

1.1. *Traditional Chinese medicine and western medicine*

Traditional Chinese Medicine (TCM) is a medical system that has been continuous practice in the past 3000 years and records thousands of herbal prescriptions, some of which are tested and validated continuously in the clinics.[3] TCM is taken by most people in China as a complementary therapeutic avenue to treat UA since its herbal remedies have some advantages over western medicine, such as fewer side effects and less cost. Historically, TCM uses pattern recognition strategies to identify patients' characteristics at individual level. These patterns, or so-called syndromes, are recognized by collecting patients' information through questioning, inspecting, and physical examinations including pulse and tongue recognitions. It is generally acknowledged that more than one distinctive clinical pattern can usually be found within each biomedical disease or condition. The clinical patterns or syndromes under each biomedical disease therefore can be viewed as unique subsets of that patient population in question. In TCM practice, specific treatment strategies including using Chinese herbal medicine or acupuncture would be applied aiming at each pattern/syndrome of a specific biomedical condition. Therefore, the research of syndromes is usually taken place in the context of a disease diagnosed by criteria established in western medicine. Thanks to clinical epidemiology and related data mining approaches, the number of syndromes and the diagnosis criteria of each syndrome in the context of many diseases have been already determined and established in China.[4,5] Nowadays, investigation material basis of syndromes in the context of biomedical diseases is a research focus, thus

building an animal model for syndromes plays a key role in understanding nature of syndromes.

1.2. *Blood stasis syndrome in TCM*

In UA patients, a syndrome named blood stasis syndrome (BSS) is found in more than 70% in the cohort.[6] It plays a key role in treating UA patients not only because it's high frequency, but because there are so many herbal remedies that have been tested effectively in clinics for treating BSS that it can help to treat UA in a personalized way.

BSS is described in TCM theory as a slowing down of the blood flow due to disruption of Heart Qi, and its main clinical manifestations in UA patients include fixed stabbing pain in the chest, aggravation of pain in the night, chest oppression and shortness of breath, palpitation, purplish tongue, and thin and unsmooth pulse. BSS is considered as a special stage of UA or myocardial ischemia in TCM and has special pathological and pathophysiologic mechanisms.

BSS is also an important underlying pathology process of many other diseases according to TCM theory. It is often understood in biomedical terms in terms of hematological disorders such as hemorrhage, congestion, thrombosis, and local ischemia (microclots) and tissue changes.

1.3. *State-of-art of animal models for BSS in the context of UA*

Due to the significant association between BSS and UA, it is urgent for the medical society to investigate material basis of BSS in the context of UA. An animal model for BSS is a better way to understand it. However, there is no available animal model with long period of stability and accurate validation for BSS since there is no widely accepted animal model for syndrome in TCM. Formally, a syndrome is clinically diagnosed by mean of four traditional phenotype information gathered methods, such as questioning, inspecting, pulse and tongue recognitions, but the related phenotype information is hard to be gathered from animals since it is impossible to inquire, listen and feel the pulse of most animals. Although inspecting information could be collected from an animal, the difference between human and an animal is large, which make symptoms have different medical meaning among clinics and animal model. Therefore, a novel strategy is needed to be presented to build and evaluate a syndrome in TCM for an animal in the context of UA.

1.4. *A computational strategy to evaluate BSS in context of UA*

The biggest obstacle of building animal model for syndrome is that it is hard to diagnose the syndrome in an animal by traditional four methods (questioning, inspecting as well as pulse and tongue recognitions) as they are performed in clinics because of the large difference of phenotype information between human and animal. However, the physical and chemical specifications of human and animal may have more

similar biomedical meaning than the phenotype information. Therefore, the novel strategy to build and evaluate an animal model for BSS is performed by two sequential steps:

1. In clinics, investigation and establishment of association between BSS and related physical and chemical specifications in the context of UA is first done computationally.
2. In animals, we build an animal model for UA and evaluate whether an animal is BSS by the association established from clinics.

Many research efforts have been performed to investigate distinct clinical biomarkers and pathological mechanisms of BSS in the past 50 years. The research strategy places heavily on discovering some specific biomarkers for BSS. However, it is found that these detected biomarkers have no specificity for BSS. Indeed, BSS (in the context of UA) is a complex disease and there are more than one biomarker. As it is pointed out that, the era of one disease with one specific gene has gone,[7] and most researchers have realized that there is no distinct single parameter of BSS caused by some kind of unbalance in regulation network, thus system biology is a better avenue to discover their biomarkers and interaction network.[3] Thus, statistical analysis methods to discover biomarkers for BSS in the system biology is turned from t test to data mining approaches. The t test is used to discover one biomarker with significant difference between case and control groups. Beyond the t test, data mining approaches can investigate significance of a pattern with more than two biomarkers between two groups. The association between BSS and physical and chemical specifications can be uncovered by data mining approaches.

In this paper, we first used data mining approaches to establish association between BSS and inflammatory factors from UA data obtained by clinical epidemiology. Alternatively, we built an animal model for UA and used the established association in clinics to evaluate whether an animal has BSS.

The paper is organized as follows. Section 2 is devoted to clinical epidemiology for UA patients. The data mining of association between syndrome and inflammatory factors is presented in Sec. 3. Building and validating an animal model for syndromes in TCM in the context of myocardial ischemia is proposed in Sec. 4. In Sec. 5 we summarize our findings and present a conclusion and discussion.

2. Clinical Epidemiology Survey for UA In-Patients

There is much evidence showing that the inflammation system is significantly associated with Coronary Heart Disease (CHD), including UA.[8,9] Furthermore, although "golden inflammatory factor" for BSS is hard to discover, it is still found that some inflammatory factors, for example, tumor necrosis factor α (TNF-α), interleukin 6 (IL-6), endothelin (ET) and nitrogen monoxide (NO), are associated with the pathology of BSS in the context of UA.[10–13] In addition, the four factors are relatively easy to be measured from animals, which make them a better bridge to

communicate between human and animal. The goal of clinical epidemiology for UA is to establish association between BSS and inflammatory factors in the context of UA.

The inclusion criteria are composed of three conditions: (1) based on diagnosis criteria of UA established in 2002 by ACC and AHA[14]; (2) patients aged between 55 and 75; and (3) patients agree to sign the informed consent. Moreover, the exclusion criteria are composed of four conditions: (1) besides UA, the patient also suffers from other cordis disease such as acute myocardial infarction, myocarditis, and cardiac nerve functional diseases; (2) a patient with angina caused by other diseases, for example, rheumatic fever, syphilis, congenital coronary anomalies, hypertrophic cardiomyopathy, cardiac mitral stenosis; (3) besides UA, a patient also suffers from stroke, diabetes, nephritis, renal failure, pulmonary infection, urinary tract infection, rheumatism, osteoarthritis, serious disease caused by liver, renal, haematogenous system, incretion system; and (4) a woman patient in gestation or lactation.

The fifty-seven UA in-patients were included in the survey in AnZhen Hospital in Beijing from September 2006 to August 2007, each of which was told to pause Resisting Platelet Activating drug for 24 hours. The 10 ml limosis vein blood of each patient was took out early morning in the next day of hospitalization. The 10 ml blood was divided into two parts. The first 2 ml was saved in the cuvette (American B-D Co.Ltd) with 2% EDTA anticoagulant and the other 8 ml was put into the counterpart with trisodium citrate anticoagulant. All samples were centrifuged with 2500 r/min for ten minutes to separate plasma from the blood, which was kept at $-80°$.

It is noted that no healthy control subjects were included since we are studying syndrome in the context of UA. BSS is used to divide the UA into two groups, UA with BSS and UA with none-BSS.

Each patient included is diagnosed whether they had BSS by TCM experts. The difference between two groups is studied by data mining approaches, which establishes the association between the syndrome and physical and chemical specifications measured in plasma.

3. Data Mining Approaches to Establish Association Between BSS and Inflammation Factors

The response variable is a categorical variable composed of BSS and non-BSS, and the independent variables are the four inflammations factors. The goal of data mining for clinical data is to build association between four inflammation factors and BSS in the context of UA.

In the first step, the independent sample t test method is employed to detect factors that are significantly different between BSS and non-BSS. Using a threshold $P < 0.01$, only TNF-α is detected that is different between the two groups. However, the students('s) t test statistics only measure the significance of single independent variable, it cannot measure significance of a pattern composed of more than two

factors between two categories. Data mining approaches provide a better solution to build association between a pattern and a response variable. They cannot only deal with data with large samples and more variables, but also for small samples. The significance of a pattern between two categories is evaluated by three performance measures of data mining approaches: accuracy, specificity, and sensitivity.

Generally, data mining methods are divided into two complementary parts. One is supervised classification and the other is unsupervised cluster. Here, the association between four factors and syndrome in TCM is investigated by classification approaches, which are mainly composed of five groups[15]: neural network, support vector machine, decision tree, Bayes approach, and logistic approach.

The accuracy of association establishment not only affects investigation of BBS in microcosmic level in clinics, but also has impact on the successful evaluation of animal model for syndrome. However, there is no evidence as to what classification approach is best fit to establish association between BSS and four factors. Consequently, we carried out a comparison study to detect the best classification approach for the UA data here. A comparison study is usually used in data mining field, especially for classification, to choose the best model for establishment of association.

We employed three hackneyed performance measures of all classification approaches: accuracy, sensitivity and specificity. A distinguished confusion matrix was obtained to calculate the three measures. A confusion matrix is a matrix representation of the classification results. The upper left cell denotes the number of samples classifies as true while they were true (i.e., true positives, TP), and lower right cell denotes the number of samples classified as false while they were actually false (i.e., true negatives, TN). The other two cells (lower left cell and upper right cell) denote the number of samples misclassified. Specifically, the lower left cell denotes the number of samples classified as false while they actually were true (i.e., false negatives, FN), and the upper right cell denotes the number of samples classified as true while they actually were false (i.e., false positives, FP). Once the confusion matrixes were constructed, the accuracy, sensitivity and specificity are easily calculated as: sensitivity $= \mathrm{TP}/(\mathrm{TP} + \mathrm{FN})$; specificity $= \mathrm{TN}/(\mathrm{TN} + \mathrm{FP})$. Accuracy $= (\mathrm{TP} + \mathrm{TN})/(\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN})$; where TP, TN, FP and FN denotes true positives, true negatives, false positives and false negatives, respectively.[16]

Three-fold cross validation was used here to minimize the bias produced by random sampling of the training and test data samples.

3.1. *Bayes approach*

Bayes classification approach originates from Bayes rules in statistics, in which Naïve Bayes and Bayesian network are two classification models that can perform the classification task.[17] We do not describe the mathematic principle of the two models in detail. Table 1 shows the results of three performance measures. The local comparison study indicates that Naïve Bayes has better performance for classification in Bayes method, so it can be taken as a 'delegate' of Bayes approach

Table 1.   The performance of Bayes in classifying BSS in the UA data.

| Approach | Models | TP | FN | Sensitivity | Specificity | Accuracy |
		FP	TN	(%)	(%)	(%)
Bayes approach	Bayesian network	41	0	100	0	71.9
		16	0			
	Naïve Bayes	40	1	97.6	37.5	80.7
		10	6			

to compete with other classification approaches. However, it is found that the better model in Bayes approach has low specificity, although the sensitivity is high, therefore, Bayes approach may be not a good way to establish association between BSS and four factors.

3.2. *Neural network*

With regard to neural network approach, multilayer perceptron (MLP) with back propagation algorithm and Radial-basis Function Network (RBF) are extensively applied in various fields for classification.[18,19] As depicted in Table 2, RBF performs better than MLP in accuracy in classifying BSS.

Both types of neural networks have hidden layers, which make neural networks have powerful classification ability. However, as the other side of coin, the hidden layers hamper the interpretation of neural network when applied to classification here, thus neural network remains a "Black Box" for us (Fig. 1). Despite of this, the powerful ability of classification by neural network still greatly helps us to establish association between BSS and inflammation factors.

3.3. *Support vector machine*

Support vector machine (SVM) was a newly developed supervised learning method during last decade (Fig. 2).[20] Here, we use two most frequently applied SVM types. One is introduced in Ref. 21, called SMO. The other is Libsvm that can be accessed in Ref. 22. Table 3 is responsible for the local comparison study results of SVM. We easily conclude that Libsvm can be considered as a "delegate" of SVM.

Table 2.   RBF is the best for neural networks.

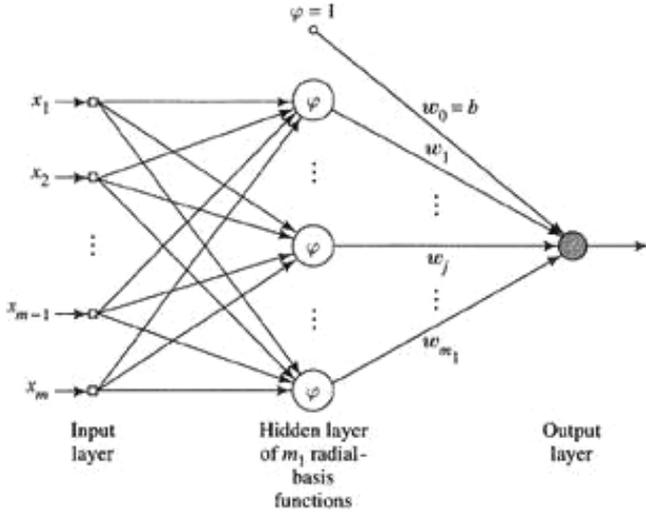| Approach | Models | TP | FN | Sensitivity | Specificity | Accuracy |
		FP	TN	(%)	(%)	(%)
Neural Networks	MLP	33	8	80.5	75	78.9
		4	12			
	RBFN	38	3	92.7	75	87.7
		4	12			

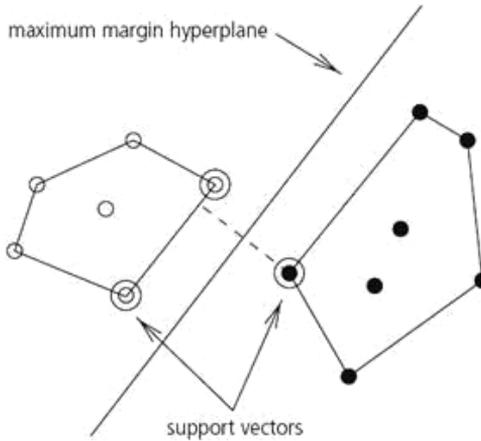Fig. 1.   The topology of MLP and RBF network.

Fig. 2.   The topology of support vector machine.

3.4. *Decision Tree*

As the name implies, this approach recursively separates observations in branches to construct a tree for the purpose of improving prediction accuracy.[23] Here we employ three kinds of decision tree classification models: J48, ADTree and Random Forest. Table 4 depicts the performance of local comparison study outcomes. The algorithm, J48 is better in the decision tree methods than other two models. The tree is illustrated in Fig. 3, from which we can see that Decision Trees is very intuitionistic, association between BSS and inflammation factors is clearer than

Table 3.   Libsvm is slightly best in a steady way.

Approach	Models	TP FP	FN TN	Sensitivity (%)	Specificity (%)	Accuracy (%)
SVM	SMO	41 16	0 0	100	0	71.9
	Libsvm	38 8	3 8	92.7	50	80.7

Table 4.   J48 is a "delegate" of decision tree.

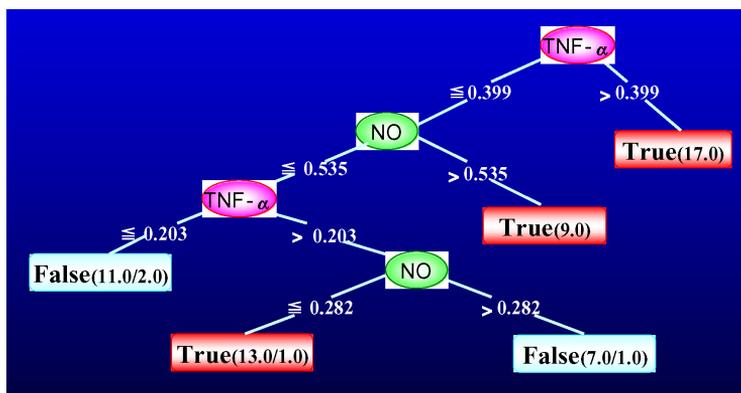| Approach	Models	TP FP	FN TN	Sensitivity (%)	Specificity (%)	Accuracy (%)
Decision Tree	J48	37 6	4 10	90.2	62.5	82.5
	ADTree	36 11	5 5	87.8	31.3	71.9
	Random Forest	36 11	5 5	87.8	31.3	71.9

Fig. 3.   The association between BSS and inflammation factors is delineated by decision tree.

other approaches. Of course, the advantage in interpretation usually goes with the slightly low accuracy in classification. It is found in Table 4 that most models in Decision Tree approach have low classification accuracy.

3.5. *Logistic regression*

Logistic regression is a generalization of linear regression.[24] It is used primarily for predicting binary or multi-class dependent variables. It contains only solo model. Table 5 shows the classification result.

Table 5.   Logistic regression performs the task well.

Approach	Models	TP FP	FN TN	Sensitivity (%)	Specificity (%)	Accuracy (%)
Logistic	Logistic Regression	36 10	5 6	87.8	37.5	73.7

3.6. *A summary of association between BSS and inflammation factors*

By global comparison, we find that radial basis function network is most fit for build association between microscopic specifications and macroscopic syndrome. The second best is J48 from decision tree approach. Although the RBF network is a kind of "Black Box" model so we could not "see" how the four factors interact with each other to behave differently between two categories, the trained RBF network is still able to predict whether an UA case is BSS.

Fortunately, the decision tree is a better approach to investigate the association. As depicted in Fig. 3, the tree is composed of two factors, TNF-α and NO. True and False in the terminal of tree represent BSS and Non-BSS respectively. From it we can see that TNF-α and NO are significantly associated with BSS and they are selected from four inflammation factors. The other two factors are considered as low association with the BSS by the decision tree approach. It is important to note that TNF-α has significant difference between BSS and Non-BSS while NO has no significance between the two groups by t test in statistics, which suggests that the data mining approaches not only take the specifications with significant difference into account, but also never "give up" specifications with no significance. Furthermore, by carefully investigation of the tree, it is found that the higher the TNF-α, the higher the possibility of a BSS diagnosis. Besides this, a lower concentration in TNF-α still induces the BSS if NO is high than 0.535 (after normalization), that is to say, BSS is associated with a combination of TNF and NO. Finally, some patients with high concentration in NO may have BSS while the other may with Non-BSS as we can see from the tree, this is why NO has no significant difference between the two groups. In a word, the association between two factors and BSS can be mined and uncovered by data mining approaches, which can help to investigate the inner mechanism of BSS.

4. Building and Evaluating Animal Model for BSS in the Context of Myocardial Ischemia

Myocardial ischemia is the nature of UA. We first built a myocardial ischemia animal model and then evaluated and predicted whether an animal is BSS by established RBF network. The animal for building model was chosen as Chinese experimental mini swine since the heart of swine is most similar to the counterpart of human as we have discussed above.

4.1. *Material*

Healthy Chinese experimental minis swine were provided by Chinese University of Agriculture, weight was 25 ± 4 kg, aged between six months and ten months, regardless of sex of swine. The healthy animals are randomly divided into two groups: model group, and sham operation group. In the former group, each Chinese experimental mini swine was instrumented with a size-matched Ameroid constrictor (Research Instrument SW, USA, inner diameter of 2.75 mm) on the anterior descending branch under general anesthesia in sterile condition. Based on early results of dynamic observations, evaluations were performed four weeks after operation. Clinical performances of animals were collected and electrocardiogram, echocardiography, coronary angiography were detected. Meanwhile, blood was obtained from former cava vena to detect blood rheological features and the concentration of prostacyclin, thromboxane A2, endothelin-1 and calcitonin gene-related peptide. Then the diseases of model animals were diagnosed.

All animals were maintained and treated in accordance with the *Principles of Laboratory Animal Care*, formulated by the National Society for Medical Research, and the g*uide for the Care and Use of Laboratory Animals*, prepared by the National Academy of Sciences and published by the National Institutes of Health (NIH Publication No. 86-23, revised 1985). The local ethics committee of Beijing University of Chinese Medicine approved all animal experiments.

General anesthesia was induced in the animals in a fasting state by intramuscular ketamine (25 mg/kg). They were intubated and anesthesia was maintained with continuous intravenous ketamine. After electrocardiogram (ECG) was performed, the heart of the animal was exposed through a left thoracotomy. A 4 to 6 mm segment of the left anterior descending coronary artery beginning at its origin was then freed by blunt dissection and an Ameroid constrictor with an internal diameter of 2.75 mm was placed around the exposed segment. The Ameroid constrictor did not interrupt the blood flow initially, but by its hygroscopic nature, it gradually occluded the vessel by external compression. After placement of the constrictor, the thoracotomy was closed by layers and the electrocardiogram was performed again.

The surgery was performed under continuous monitoring of ECG. After recovered from anesthesia, animals were extubated. During the first three days after surgery, penicillin (4,800,000 units per day) was injected intramuscular to anti-infection.

4.2. *Evaluating the MI disease of animal model*

Selective coronary angiography was performed after four weeks and the degree of narrowing of the anterior descending branch were observed.

Echocardiography study was conducted pre and four weeks post surgery respectively to evaluate the degree of stenosis and myocardial function. Echocardiography studies including the images of six standard planes (parasternal long axis, short axis at bicuspid level, papillary muscle level, cardiac apex level, apical four chambers and

apical two chambers), adopting LVEDd, LVEDs, IVST, LVAW thickness (bicuspid level, papillary muscle level, cardiac apex level), EDV, ESV, peak early diastolic velocity, peak late diastolic velocity, etc, and calculating EF, FS, SV, $\triangle$T. 2-DE is introduced to visual examination the movement of the left ventricular muscles.

Once the animal model for MI is built, by using the established association, each animal can be evaluated whether is with BSS based on the "communication bridge" — the four inflammation factors. However, the difference of the four factors in concentration contained in the blood is needed to be investigated before evaluation. That is to say, the four factors should have the same scale on human and swine.

4.3. *The significant difference of inflammation factors between animal and human*

We found that each inflammation factor takes different concentration between animal and human. The student's t tests showed that there is a significant difference between microscopic specifications of animal and human (Table 6). In order to bridge the gap, we employed normalization method for each group to scale the concentrations of each factor. The normalized method was given in Eq. (1):

$$\overline{X} = \frac{X - \min(A)}{\max(A) - \min(A)}, \tag{1}$$

where X represents factors, $\overline{X}$ is responsible for the normalized factors, A is the group (human or animal) where X is included, and $\min(A)$ is the minimal value of X in group A while $\max(A)$ is the maximal value of X. It is easy to see that $\overline{X}$ values between 0 and 1.

As shown in Fig. 4, after normalization with regard to each group, the concentrations of factors are in the same scale, which paves a basis for further prediction. Otherwise, a great error would be occurred during the prediction of whether an MI animal is BSS.

4.4. *Evaluating whether an animal is BSS by RBF network in the context of MI*

As shown above, human data was used to train the RBF network and animal data was used for test animal model by using the model built by RBF network. The RBF network has the ability to predict whether a new case is BSS after training

Table 6.   Mean and standard derivation of each factor for human and animal respectively.

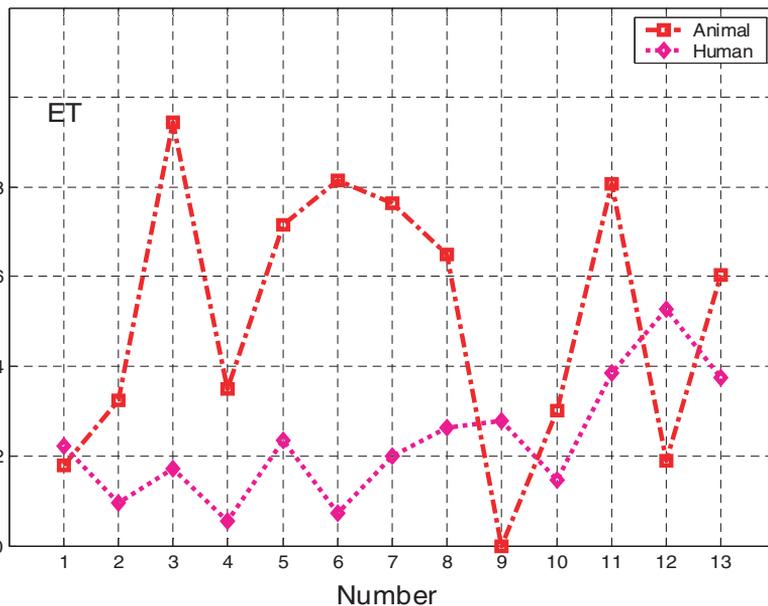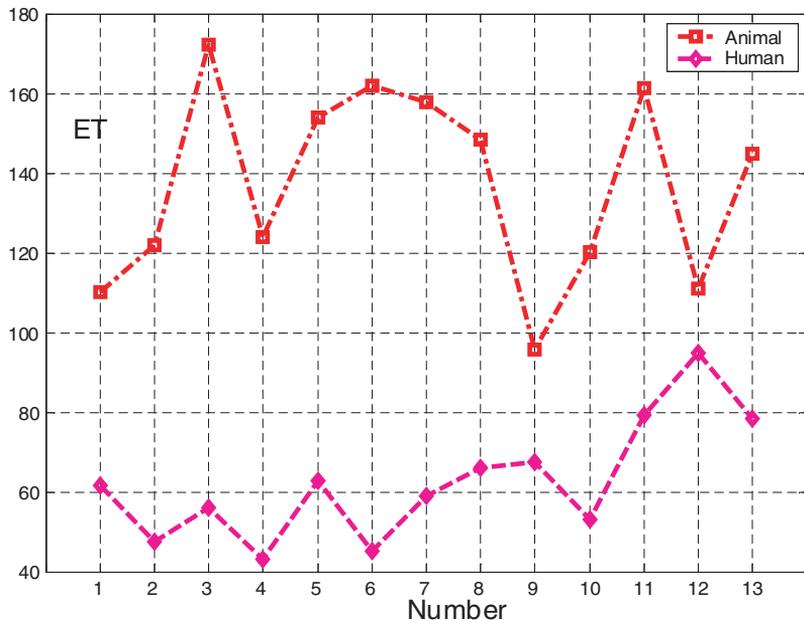|	ET	TNF	IL-6	No
Human	66.611 ± 18.576	0.68053 ± 0.24097	68.94 ± 47.134	82.912 ± 54.451
Animal	137.16 ± 24.391	1.7656 ± 0.53346	320.39 ± 98.939	57.747 ± 29.497

Fig. 4. Thirteen patients that are randomly chosen from the cohort are compared with 13 swine. The above figure shows that the minimum of ET concentration is larger than maximal of counterpart in human, which means that the difference of four inflammation factors between animal (swine) and human is significant. After normalization with regard to each group, as depicted in nether figure, the factors are in the same scale between zero and one.

Table 7.   Parameter setting for RBF network.

Parameter	Description	Setting
Clustering Seed	The random seed to pass on to K-means	1
Ridge	Set the Ridge value for the logistic or linear regression	$1.0E - 8$
Maxiam Iteration	Maximum number of iterations for the logistic regression to perform	-1
minStdDev	Sets the minimum standard deviation for the clusters	0.1
numClusters	The number of clusters for K-Means to generate	2

Table 8.   The prediction of whether a disease animal is BSS or non-BSS by using two better classification approaches. It is noted that the normalization of animal model is with regard to all animals, including sham operation groups (data not shown here since they are not needed to be evaluated by the approaches).

Animal	ET	TNF-	IL-6	NO	Prediction by RBF	Prediction by J48
1	0.17948	1	0.52407	0.11825	FALSE	TRUE
2	0.32296	0.13656	0.37839	0.38443	FALSE	FALSE
3	0.94394	0.77967	0.92968	0.49553	TRUE	TRUE
4	0.34929	0.67016	0.34369	0.68466	TRUE	TRUE
5	0.71667	0.5637	0.20881	0.36932	TRUE	TRUE
6	0.81531	0.30977	0.78601	0.30622	TRUE	FALSE
7	0.76521	0.85297	0.73035	0.43243	TRUE	TRUE
8	0.64986	0.52443	0.78601	1	TRUE	TRUE
9	0	0.51134	0.6987	0.49499	TRUE	TRUE
10	0.30225	0.66274	0.64863	0	FALSE	TRUE
11	0.80805	0.6514	0.48062	0.18028	TRUE	TRUE
12	0.1889	0.363	0.24321	0.11709	TRUE	TRUE
13	0.60479	0.45899	0.91622	0.36986	TRUE	TRUE

by the clinical data. The parameter information for RBF network that configures the classification model is shown detailed in Table 7. Each animal with myocardial ischemia disease can be predicted by the network. The detailed results are shown in Table 8. We can see that ten animals from a total of 13 animals are with BSS, the other animals are with non-BSS, which means that BSS in the context of MI is a subtype of MI (UA). By concept of syndrome in TCM, MI cohort can be divided into different groups, each of which has same phenotype that is characterized by the syndrome. Chinese herbal is prescribed in accordance with syndrome diagnosed. Therefore, personalized therapy of UA should be taken into account. Furthermore, the frequency of BSS in the context of animal model is nearly 77%. While in the clinics, the frequency of BSS in the context of UA is about 72%, the frequency error is 5%; which can be considered as sampling error since there are only 13 animals to be included here. It indicated that the evaluation of BSS in the animal model is accurate to some extent. Moreover, we used the second best classification approach (Decision tree J48) to re-evaluate again the 13 animals, we found out that most of animals are evaluated as same kind of syndrome. So the results obtained by RBF network evaluation and prediction are robust and credible.

5. Conclusion and Discussion

In this paper, we proposed a novel strategy to build and evaluate an animal model for BSS in TCM in the context of UA. The work filled the blank of adequately evaluating animal model for syndromes and solved the problem of how to diagnose syndromes in animal. We took advantage of supervised data mining approaches to establish the association between physical and chemical specifications and a syndrome in the context of the disease in clinical data obtained by clinical epidemiology survey. The accuracy of classification of data mining approach guarantees the association establishment is right (higher than 87%), then the specifications were used as "Communication Bridge" to translate the association to the animal. The prediction results showed that an animal model with the same disease (UA) may have different syndrome phenotypes and the association established in the clinics could be used to evaluate whether an animal has BSS. The presented strategy here not only builds and evaluates an animal model for syndrome in TCM, but also paves a key basis to uncover the mechanism of syndromes and treat disease in a personalized way.

The paper only took the most important syndrome in the UA-BSS into account. So the UA cohort is divided into two subgroups: BSS and non-BSS. However, through prior clinical research, it was found that there are about seven syndromes can be discovered in the context of UA. Further work will focus in the differential diagnosis of each syndrome by physical and chemical specifications in the context of UA and build animal models for them respectively.

Acknowledgments

References

1. Monnet E, Chachques JC, Animal models of heart failure: what is new? *Ann Thorac Surg* **79**:1445–1453, 2005.
2. Weaver ME, Pantely GA, Bristow JD, Ladley HD, A quantitative study of the anatomy and distribution of coronary arteries in swine in comparison with other animals and man, *Cardiovasc Res* **20**:907–917, 1986.
3. Li S, Zhang X, Li Y, Wang Y *et al.*, Understanding Zheng in traditional Chinese medicine in the context of neuro-endocrine immune network, *IET Syst Biol* **1**(1):51–60, 2007.
4. State Food and Drug Administration, *Guiding Principle of New Drug Clinical Research of Traditional Chinese Medicine*, Medical Science Publishing House of China, Beijing, 2002 (Chinese).

5. Chen J, Xi G, Chen J, Zhen Y, Xing Y, Wang J, Wang W, An unsupervised pattern (syndrome in traditional Chinese medicine) discovery algorithm based on association delineated by revised mutual information in chronic renal failure data, *J Biol Syst* **15**(4):435–451, 2007.

6. Sun Z, Xi G, Yi J, Zhao D, Select informative symptom combination of diagnosing syndrome, *J Biol Syst* **15**(1):27–37, 2007.

7. Hayden EC, Biological tools revamp disease classification, *Nature* **453**(7156): 709, 2008.

8. Gustavsson CG, Agardh CD. Inflammatory activity increases with haemoglobin A_1c in patients with acute coronary syndrome, *Scand Cardiovasc J* **16**:1–6, 2009.

9. Davidson KW, Schwartz JE, Kirkland SA *et al.*, Relation of inflammation to depression and incident coronary heart disease (from the Canadian Nova Scotia Health Survey [NSHS95] Prospective Population Study), *Am J Cardiol* **103**(6):755–761, 2009.

10. Mao Y-L, Yuan Z-K, Huang X-P, *et al.*, Study on relationship between the polymorphism of angiotensin converting enzyme gene and blood stasis syndrome in patients with coronary heart disease, *Chin J Integr Tradit West Med* **24**(9):776–780, 2004 (Chinese).

11. Yuan Z-K, Huang X-P, Tan G-B, *et al.*, Analysis of the function of vascular endothelial cells in coronary heart disease patients of blood-stasis syndrome, *Chin J Integr Tradit West Med* **26**(5):407–410, 2006 (Chinese).

12. Zhang J-G, Yang N, He H, *et al.*, Effect of astragalus injection on plasma levels of apoptosis-related factors in aged patients with chronic heart failure, *Chin J Integr Tradit West Med* **25**(11):187–190, 2005 (Chinese).

13. Ma X, Yin H, Chen K, Research progress of correlation between blood stasis syndrome and inflammation, *Chin J Integr Tradit West Med* **27**(7):669–672, 2007 (Chinese).

14. Braunwald E *et al.*, ACC/AHA 2002 guideline update for the management of patients with unstable angina and non–ST-segment elevation myocardial infarction — summary article. A report of the American College of Cardiology. American Heart Association Task Force on Practice Guidelines (Committee on the Management of Patients with Unstable Angina), *J Am Coll Cardiol* **40**(7):1366–1374, 2002.

15. Witten IH, Frank E, *Data Mining: Practical Machine Learning Tools and Techniques*, 2nd ed., Morgan Kaufmann, 2005.

16. Delen D, Walker G, Kadam A, Predicting breast cancer survivability: a comparison of three data mining methods, *Artif. Intell. Med.* **34**(2):113–127, 2005.

17. Morelande M, Kreucher C, Kastella K, A Bayesian approach to multiple target detection and tracking, *IEEE Trans Signal Process* **55**(5):1589–1604, 2007.

18. Lee K, MLP-based phone boundary refining for a TTS database, *IEEE Trans Audio Speech* **14**:981–989, 2006.

19. Peng J, Li K, Irwin G, A novel continuous forward algorithm for RBF neural modeling, *IEEE Trans Automat Contr* **52**:117–122, 2007.

20. Vapnik V, *The Nature of Statistical Learning Theory*, Springer, New York, 1995.

21. Platt J, Training of support vector machines using sequential minimal optimization, in Burges C, Smola A (eds.), *Advances in Kernel Methods — Support Vector Learning*, MIT Press, 1980.

22. Chang C, Lin C, LIBSVM: A library for support vector machines, 2001. Software available at http://www.csie.ntu.edu.tw/∼cjlin/libsvm.6.

23. Shekhar R, Phoha G, Balagani K, K-Means+ID3: A novel method for supervised anomaly detection by cascading K-Means clustering and ID3 decision tree learning methods, *IEEE Trans Knowl Data Eng* **19**:345–354, 2007.

24. Hastie T, Tibshirani R, Friedman J, *The Elements of Statistical Learning*, Springer-Verlag, New York, 2001.