Computational methods for Traditional Chinese Medicine: A survey

Suryani Lukman[a], Yulan He[b,*], Siu-Cheung Hui[c]

[a] *Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK*
[b] *Informatics Research Centre, University of Reading, Reading RG6 6BX, UK*
[c] *School of Computer Engineering, Nanyang Technological University, Nanyang Avenue, Singapore 639798, Singapore*

ARTICLE INFO

ABSTRACT

Traditional Chinese Medicine (TCM) has been actively researched through various approaches, including computational techniques. A review on basic elements of TCM is provided to illuminate various challenges and progresses in its study using computational methods. Information on various TCM formulations, in particular resources on databases of TCM formulations and their integration to Western medicine, are analyzed in several facets, such as TCM classifications, types of databases, and mining tools. Aspects of computational TCM diagnosis, namely inspection, auscultation, pulse analysis as well as TCM expert systems are reviewed in term of their benefits and drawbacks. Various approaches on exploring relationships among TCM components and finding genes/proteins relating to TCM symptom complex are also studied. This survey provides a summary on the advance of computational approaches for TCM and will be useful for future knowledge discovery in this area.

1. Introduction

Traditional Chinese Medicine (TCM) is a range of medical practices used in China for more than four millenniums, a treasure of Chinese people. TCM has also been recognized as a popular complementary and alternative medicine in Western countries because Chinese medicine is generally extracted from natural products without artificial additives which creates mild healing effects and incurs fewer side effects. TCM theory is based on philosophical frameworks such as the Theory of Yin-Yang and Five Elements, the human body meridian systems and the Zang Fu Theory. According to holistic viewpoint of TCM, the human body is constantly interacting with the environment, hence combinatorial signs of disharmony in the external and internal environment of a person are considered in understanding, preventing, and treating diseases. TCM diagnostics and treatments are popular in East Asia as their costs are much lower than Western approaches which require sophisticated equipment or extensive chemical processing.

Computational methods for TCM allow researchers to identify required information more efficiently, discover new relationships which are obscured by merely focusing on Western medicine, and bridge the gaps between Western Medicine and TCM. Data mining, including text mining and knowledge discovery, promises a solution for dealing with information overloading. Data mining techniques have been used to explore active ingredients in the effective TCM formulations for individual diseases. Moreover, by creating formalized knowledge for expert systems, a second diagnostic opinion may be obtained.

With the increasing acceptance of TCM in Western medical field, it is important to highlight their differences and similar-

ities. In this way, biomedical mining of TCM can be focused on problems addressable by TCM but not by Western medicine. Modern Western medicine attempts to identify the individual factors contributing to illness, such as bacteria and viruses. On the other hand, TCM views the causes of illness as symptoms of diseases. Diseases are regarded as the result of abnormal interactions or imbalances in human system. This can be explained in a way that if a particular organ is unhealthy, it will be vulnerable to external invasion such as heat, cold, etc. Since functional imbalance and specific manifestations of disease are differentiated as syndrome, the concept of syndrome differentiation is important in the TCM diagnostics [1].

In TCM formulations, the balance and interaction of all the components are regarded as more important than the effect of individual component, which is in contrast to the principle of Western medicine. This is because other components may be added to the TCM formulation to suit the patient's Yin and Yang conditions or to nullify the toxicity or side effects of the main components. Nonetheless, there are also similarities between the Western medicine and TCM. For example, the interactions between the clinically active molecules of Western medicine or TCM and their biological targets at the molecular level are the foundation of disease treatment.

This study is aimed to survey the TCM data sources available *hitherto* and to illustrate the progress of biomedical mining for TCM using computational approaches. A brief presentation on TCM elements is included to assist readers in understanding the rationales of various computational techniques for TCM. The foci of numerous computational approaches reviewed herein address issues in TCM formulations and TCM diagnosis. Knowledge base grid applications in TCM are also discussed as they are highly desirable for future research on TCM. Finally, the paper concludes with a discussion on future trends.

2. Elements of TCM

In TCM, human body is balanced by two opposite cosmological forces, they are Yin and Yang as shown in Fig. 1. Disturbance of Yin results in internal dysfunction, whereas Yang is the external causing agents including climatic, infectious, and contagious. The relationships between Yin and Yang are typified by the Five Elements, which are Wood, Fire, Earth, Metal, and Water. Each of these five elements, interacting and counterbalancing one another, represents respective aspects of TCM, namely organs, tastes, human sounds, emotions, etc. In accordance to the Zang Fu Theory, heart, liver, spleen, lung, kidney, and pericardium are classified as the Zang organs, whereas the Fu organs comprise small intestine, large intestine, stomach, gall bladder, and urinary bladder. A brief knowledge on these is required for understanding the rationales and various approaches of TCM, consisting of TCM formulations, diagnostic processes, and treatments.

TCM formulations generally consist of complex mixtures of several, many times more than 20 various herbal plant parts and sometimes animal parts. In fact, almost 83% of 5500 natural sources of TCM listed in the last encyclopedia (1977) are derived from plants [2]. Herbal plants contain plethora of natu-

ral products which are abundant sources of active compounds for treating diseases. The advantage of natural products over synthetic compounds lies in their inherently large-scale structural diversity.

TCM diagnostic processes are based on four methods: inspection, auscultation and olfaction, inquiring, and palpation [3]. In the *inspection* approach, TCM practitioners observe abnormal changes in the patient's vitality, color, appearance, secretions and excretions. The vital signs encompass eyes, tongue, facial expressions, general and body surface appearance. The inter-relationship between the external part of the body such as face and tongue and the internal organ(s), is used to assist TCM practitioners to predict the pathological changes of internal organs. *Auscultation* refers to listening of the patient's voice, breathing, and coughing and is used to judge the pathological changes in the interior of the patient's body, whereas *olfaction* refers to smelling of secretion or excretion products. *Inquiring* or *interrogating* is to query patient's family history, feelings in various aspects, such as chills and fever, perspiration, appetite and thirst, as well as pain in term of its nature and locality. *Palpation* approach involves pulse diagnosis.

The TCM treatment techniques include acupuncture and moxibustion, massage therapy, Qigong, surgery, breathing exercise, and TCM formulations.

This survey focuses on the computational approaches to TCM formulation and diagnosis processes including inspection, auscultation and olfaction, inquiring, and palpation.

3. TCM herbs and formulations

TCM formulations can be classified in accordance to several categories, namely The Four Natures, The Five Tastes and The Channel Tropisms[1]. The Four Natures are related to the degree of Yin and Yang, ranging from cold (extreme Yin), cool, neutral semi-warm to hot (extreme Yang). The Five Tastes of TCM formulations are pungent, sweet, sour, bitter, and salty. The of Channel Tropism refer to the target organs. In this section, we provide a list of various databases useful for studying TCM herbs and formulations. We also discuss a significant challenge in TCM herbs and formulations mining, named entity recognition, followed by numerous data and text mining tools developed so far.

3.1. Databases

To answer the challenge of acceptance and development of TCM due to the lack of comprehensive understanding on TCM progression and disease treatment at the molecular level, thorough data on TCM formulations are required. An ideal database should allow efficient data mining. This section attempts to analyze a number of TCM formulation databases in the context of overall usability and, if available, integration with other databases or approaches such as Western medicine. A list of TCM formulation databases is given in Table 1. They were chosen mainly due to their popularity and
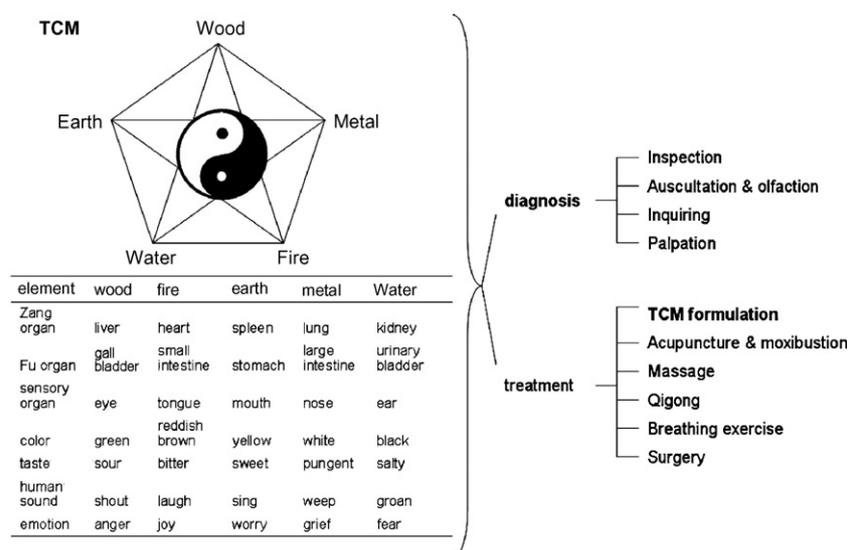
---

[1] http://www.51qe.cn/pic/30/11/76/022.htm.

Fig. 1 – The relationships among various TCM elements.

Table 1 – Databases of TCM herbs and formulation		
Databases	Descriptions	URL or references
Acupuncture.com.au	Classification of TCM formulations based on their actions	http://www.acupuncture.com.au/education/herbs/herbs.html
Dictionary of Chinese Herbs	Disease-specific TCM formulations, toxicity and side effects	http://alternativehealing.org/Chinese_herbs_dictionary.htm
Plants For a Future	TCM herbs with their potential side effects, physical characteristics, and medicinal usages substantiated by relevant scientific citations	http://www.pfaf.org
3D structural database of biochemical components	TCM herbs with basic molecular properties, optimized 3D structures, the herbal origin with English and Latin names, and clinical effects	Qiao et al. [4,5]
Traditional Chinese Medicine Database (TCMD)	Information on Chinese medicines, original plants and bioactive compounds	He et al. [6]
TCM Knowledge Base Grid	TCM medicine database, traditional Chinese drug database, TCM literature databases, traditional Tibetan drug database	http://www.cintcm.com
TCM drugs information system	Chinese herb database, Chinese patent medicine database, effective components database of Chinese herbs, Chinese medical dietotherapy prescription database, and Chinese medical recipe database	Qiao et al. [4]

the evidence found in the literature that various data mining tasks have already been successfully performed on them.

Acupuncture.com.au provides a list of TCM formulations grouped in accordance to their actions. Inclusion of both the English and Chinese names of TCM herbs is useful in cross-studying TCM herbs using traditional and modern experimentations. Moreover, their tastes and contraindications are also recorded.

The *Dictionary of Chinese Herbs* provides information on the toxicity and side effects of TCM herbs. It also includes treatments and samples of TCM formulas for diseases such as cancer, dengue fever, diabetes, and hepatitis B. A compilation on TCM herbs which are not compatible with certain drugs, containing 66 pairs of incompatible TCM herbs and drugs as of 6 November 2006, is also included. This provides biochemical

explanations for the incompatibility of TCM herbs and drugs to the drug designers.

The *Plants For a Future* database allows retrieving TCM herbs with specific medicinal usage. For example, a query on *Rehmannia glutinosa* returns information on its common/Romanized Chinese name,[2] potential side effects, physical characteristics, and medicinal usages substantiated by relevant scientific citations.

A *3D structural database* of the biochemical components extracted from TCM herbs is also available [4,5]. The active compounds discovered from TCM formulations can be used as

---

2 Search can also be performed using common names in addition to Latin names.

probes to identify treatment target and proteomic functions. The database comprises 10,564 herbal component records and 2073 TCM herbs entries of 296 families. Detailed information such as basic molecular properties, optimized 3D structures, the herbal origin with English and Latin name, and clinical effects are included. A potential usage of such a 3D database includes design of new drug through finding the common structural characteristics among bioactive components extracted from TCM herbs with same curative effects on a particular disease.

The *Traditional Chinese Medicine Database* (TCMD) [6] was constructed mainly from the book "Traditional Chinese Medicines: Molecular Structures, Natural Sources and Applications" [7]. The authors attempted to relate TCM to Western medicine. The overall TCMD organization comprises Chinese medicines, original plants, and bioactive compounds. It stores more than 1540 species of herbs or natural products used as TCM and 6800 molecular compounds isolated from plants, herbs, animals and fungi.

There are other databases not specifically for TCM herbs or formulations, but it is possible to retrieve relevant information on TCM herbs or formulations from them. A list of online TCM databases can also be found in Ref. [8]. Besides those databases, the Information Institute of TCM Academy[3] has provided several large-scale databases used in TCM Knowledge Base Grid [9] such as patient record database, TCM medicine database, traditional Chinese drug database, TCM literature databases, traditional Tibetan drug database, etc. TCM Knowledge Base Grid will be discussed in Section 6. A TCM drugs information system based on networks of five large databases has also been developed [4]. It includes Chinese herb database, Chinese patent medicine database, effective components database of Chinese herbs, Chinese medical dietotherapy prescription database, and Chinese medical recipe database.

3.2. *Named entity recognition (NER)*

The concept of a named entity was first introduced at the Message Understanding Conference (MUC) [10]. Named entity defines some informative words, such as disease names, organ names, as information units in text. Recognition of named entity is aimed to further extract relationships and other information by identifying the key ideas or concepts. Approaches to NER include dictionary-based (lexicon-based), rules-based, machine learning and statistical approaches, and hybrid approaches.

The word segmentation problem in Chinese makes the NER problem harder as an error in the word segmentation step may lead to errors in NER results [11]. The char-based NER, which bypasses the word segmentation step [12,13], may be able to solve the problem. Although there exists a Unified Traditional Chinese Medical Language System (UTCMLS), which provides large-scale ontology for TCM entities [14], simple text-matching algorithm is still insufficient to fulfill the NER task. The same word, phrase, and expression may have multiple context-dependent meanings. To deal with

one-term-multiple-concepts, the term from alphabet can be eliminated and only the expressions for detailed concepts are reserved [15].

On the other hand, many TCM entities have several names. This is because the fundamental principles of TCM are based on several schools of thought, which are not necessarily uniform. To deal with the multiple-terms-one-concept case, Li et al. [15] only reserve the definition for the uniformed concept. The *standard syndrome-name database* aims to deal with the problem that a syndrome was assigned with different names by different TCM practitioners [1]. Generally, a standard syndrome name always includes minimal a key element indicating the disease occurrence site (such as lung and heart) and another key element on disease causes or pathological causes (such as wind, cold, dampness, etc.). As there is no standardized guidance for translation of TCM entities from Chinese to English, NER is further complicated. Some traditional terms, definitions or expressions have resulted in semantic ambiguities, hence efficient techniques for disambiguation need to be studied.

3.3. *Mining tools for TCM herbs and formulations*

The currently available data mining tools for TCM herbs and formulations are insufficient for addressing the challenges of overgrowing TCM data in numerous databases. Table 2 summarizes some of the data mining tools for TCM formulations.

A text mining system MeDisco/3T (Medical Discover for Traditional Treatment inTelligence) was developed to mine the TCM herbal knowledge from TCM literature [16]. MeDisco/3T extracts structured TCM formulation information such as TCM herbal names, Chinese herbal medicine components, and efficacy description from literature based on the bootstrapping approach which iteratively produces new seeds and patterns using a small set of initial seeds. The average name extraction precision using bootstrapping achieves over 95%. After that, the association rule mining algorithm, Apriori [19], is used to analyze the TCM formulation data, in which the frequent Chinese herbal medicine pairs and Chinese herbal medicine family combinations are found. The principle used for classification is due to the fact that all the clinical TCM formulations used in TCM can be classified by its efficacy. More than 18,000 TCM formulations are stored and studies on frequent herbal pairings have been conducted.

Cao et al. [17] developed an ontology-based system for extracting knowledge on TCM drugs and formulations from semi-structured text. Different from named entities which define some informative words, such as disease names, organ names, as information units in text, an ontology is a specification of the concepts and relationships. The system comprises two parts: an executable knowledge extraction language (EKEL) for specifying knowledge-extracting agents and a support machine for executing EKEL programs. They also developed a TCM formulation ontology from a list of textbooks, codices, encyclopedias and dictionaries. Using their system, knowledge of more than 2710 traditional Chinese drugs and 5900 traditional Chinese formulations can be obtained.

Zhang et al. [18] proposed a compound innovation method which integrates the possibility construction space theory and

---

Table 2 – Data mining tools for TCM formulations and their approaches.		
System	Approaches	References
MeDisco/3T	TCM-related information extraction based on bootstrapping and frequent itemset mining using the Apriori algorithm	Zhou et al. [16]
Ontology learning system	Extracting knowledge on TCM drugs and formulations from semi-structured text	Cao et al. [17]
Compound innovation method	Possibility construction space theory and TCM formulation theory are integrated to design a novel TCM compound	Zhang et al. [18]
TCMiner	Frequent pattern mining, association rules generation, frequent closed patterns and rules mining	Li et al. [15]

the TCM formulation theory in order to design a novel TCM compound. Prior to the application of the compound innovation method, similarities between two individual drugs need to be measured based on the TCM formulation characteristics, such as their natures and tastes, as well as the drug targets and functions.

TCMiner is a data mining system developed using Delphi to analyze TCM prescriptions [15]. The implemented algorithms include frequent pattern mining, association rules generation, and frequent closed patterns and rules mining. In their performance study, the authors claimed that their frequent patterns mining and association rules generation algorithms outperformed FP-Growth and Apriori, respectively [20].

4. TCM diagnosis

As mentioned earlier in Section 2, TCM diagnosis generally comprises four approaches: inspection, auscultation and olfaction, inquiring, and palpation. Since the results of TCM diagnosis may vary among individuals due to different experience and environmental factors [21], a computational approach for assisting TCM diagnosis will be much sought after. We focus our review on computational methods developed for inspection, auscultation, and pulse analysis, which is a part of palpation. Furthermore, we also discuss expert systems for TCM diagnosis, which mostly encompasses the four aspects of TCM diagnostics.

4.1. Inspection

In the diagnosis through inspection, a patient's tongue color is among those features traditionally observed. According to TCM principles, different portions of tongue reveal the health information of the corresponding internal organs of a human body. For example, the tip of the tongue reveals the heart and lung conditions whereas the middle tongue discloses the conditions of the spleen and stomach. Tongue image analysis has also been used to evaluate the therapeutic effects of TCM formulations [22]. Attempts to develop computational visualization techniques for tongue analysis are still at an early stage of development, but works have been in progress. The relationship between some diseases and abnormalities in the patient's tongue and tongue coating has been substantiated by clinical evidence, for example the association between the various viscera cancers and the changes of color, coating, degree of wetness and coarseness, shape and dorsum shape

of the tongues [23], has motivated the development of various tongue visualization techniques.

There are several issues to be addressed in computerized tongue image analysis. The methods to capture the tongue image, followed by effective segmentations of the tongues and calibrations of their colors need to be considered. Generally, two kinds of quantitative features, chromatic and textural measures, are extracted from tongue images. Selection and representation of tongue image features are the next issues to be tackled with. In the early studies, charge-coupled device (CCD) cameras were used to capture the tongue images and an red, green, blue (RGB) model was used to represent the colors [24]. The next problem is the fuzzy characteristics of the data itself. The more recent approaches used so far to address these problems include fuzzy theory [25], neural networks [26], parallel coordinate visualization [23], Bayesian networks [27,28], and integrations of them.

4.1.1. Tongue property identification
Since the color property alone is insufficient to reveal much information on the tongue coating, a computerized tongue examination system (CTES) exploiting chromatic and textural algorithm was developed to quantize the tongue properties [24]. The chromatic algorithm, which is based on the hue, saturation, luminance (HSL) color model, serves to identify the colors of the tongue and the thickness of its coating, whereas the textural algorithm functions in detecting the grimy coating. CTES identifies five colors (i.e. whitish, light red, regular red, maroon, and purple) from the substance and three colors (i.e. white, light yellow, and yellow) from the coating. The formulae of some spatial gray-tone dependency matrices (SGTDM)-based textural features such as angular second moment (ASM), contrast, correlation, variance and entropy are used to determine the textural-related property of tongue. The author reported that the overall classification correctness of the system exceeded 86%. The limitation of this study is the difficulty of identifying the properties of wetness and dryness on the tongue by the computer, which can be easily done by TCM practitioners.

4.1.2. Segmentation
Simple segmentation methods are limited by the weak edge of tongue and the details on the tongue's surface, hence a unique segmentation method based on the combination of the watershed transform and an active contour model (ACM) was proposed [29]. The watershed transform is used to get the initial contour whereas the active contour model is used to

converge to the exact edge. The watershed transform treats an image containing a tongue as a surface where dark regions are low and light regions are high, then the transform finds watershed ridge lines between high and low regions and catchment basins in low regions. The tongue region generally corresponds to the light region. Next, the curves of tongue image are drawn toward the edges and kept smooth and continuous by means of energy minimization in the parametric active contour model. The authors reported that 96.7% of the 120 tongue image were segmented properly. Nonetheless, the more details on the surface of the tongue, the harder the segmentation process will be.

A research utilizing JSEG, an approach for unsupervised segmentation in two independent stages of color quantization and spatial segmentation [30], was employed to partition the tongue into several homogenous areas [31]. This was followed by applying a priori template for detecting tongue substance and coat. Next, color classification was performed by a nearest neighbor classifier. Based on 600 samples, their color separation results achieved more than 90% agreement with TCM practitioners' opinion, but the authors did not perform any significance test on the accuracy of their method.

4.1.3. Diagnosis on tongue inspection

A systematic diagnostic supporting system of tongue inspection using several logic scores based on fuzzy theory was developed to diagnose syndromes such as deficiency syndrome, excess syndrome, coldness syndrome, and heat syndrome [25]. It mainly focuses on coldness and heat syndromes in which five grades of color (pale white, pale, light red, reddened, and deep red), five grades of color of fur (white, slight yellow, yellow, dark yellow, and black), and five grades of wetness–dryness are used. On the other hand, for analyzing the condition of deficiency and excess syndromes, the shape of the tongue and the thickness of fur are employed. For each condition, seven grades of diagnostic outcomes are labeled with input from TCM experts (e.g. seriously cold, moderately cold, slightly cold, normal, slightly hot, moderately hot, seriously hot). Fuzzy rules are then constructed to connect different diagnostic labels (input) and tongue diagnostic outcomes (output). The relevance between tongue inspection by TCM experts and their systems derived from those fuzzy rules are verified on 20 healthy volunteers for coldness and heat syndromes. Their results demonstrated 90% coincidences. However, there is a lack of system testing on sick patients and it is hard to tell how accurate the system would be on those sick patients' data.

A multi-layered feed forward neural network and error back-propagation learning rule with RGB histograms as inputs were used to analyze the patients' tongue colors [26]. Features used include tongue color (light pink, rose pink, red, purple or blue), tongue thickness, tongue shape (round, triangle), fur color (white, yellow, black), fur thickness, degree of wetness, and fur shape. Assuming tongue color is the most crucial factor, their neural network performance for analyzing the tongue color ranges from 16.7% to 77.8%. Similar to the previous work by Watsuji et al. [25], the approach only tested on healthy patients.

Parallel coordinate visualization and integration of fuzzy sets were used to classify five health states which reflect can-

cers, based on six features extracted from a set of 34 tongue images [23]. The five health states are healthy, history of cancers, history of Polyps, Polyps, and Colon Cancer. The six features are texture features related to energy and entropy, chromatic dimensions of a* and b* color space, the roughness and the amount of cracks on the tongue. The parallel coordinate approach is capable to deal with that which other generally used clustering and classification methods such as K-means, C4.5, SVM and neural networks failed to achieve. Additionally, this study has provided a platform for future work of a complicated visualization by integrating fuzzy sets and the three-dimensional parallel coordinate approach when a much larger dataset is available.

A Bayesian network is a causal probabilistic network that compactly represents a joint probability distribution (JPD) among variables of interest. Discrete Bayesian networks were employed to model the relationship between the quantitative chromatic cum textural features and diseases [27]. In addition to the commonly used RGB color space, they also employed CIEYxy, CIELUV, and CIELAB color spaces [32], for the extraction of quantitative color features. The effectiveness of the method was evaluated on a group of 455 patients affected by 13 common diseases as well as other 70 healthy volunteers, and the accuracy of diagnostic results predicted by the previously trained Bayesian network classifiers achieved 75.8%. Nevertheless, the method was only compared with the nearest neighbor classifier. Bayesian networks were also used to establish the relationship between quantitative chromatic-textural features and diseases [28]. The system could identify six groups: healthy, pulmonary heart disease, appendicitis, gastritis, pancreatitis and bronchitis with accuracy higher than 75% of the 544 tongue images tested. Nevertheless, four groups of diseases were poorly classified.

4.2. Auscultation

According to the TCM theory, states of Ying and Yang which correspond to changes in the viscera function, may lead to changes in voice. Computational methods to address auscultation in TCM diagnosis are scarce at the moment due to the lack of objective assessment and accurate quantification. Nonetheless, a few studies have been done in this area. The rationale is any malfunction of an internal organ would change the pattern of voice, which is detectable through pattern analysis.

Acoustic parameters of voice were proposed in order to provide a quantitative analysis of auscultation in TCM diagnosis [33]. It was discovered that the most commonly used voice parameters, pitch frequency or pitch variations, the power of the voice, and the speed of speech, are useless in analyzing the changes in patients' voice. Among the four proposed parameters of the average number of zero-crossings, the variations in local peaks and valleys, the variations in first and second formant frequencies, and the spectral energy ratio, two temporal parameters, the average number of zero-crossings and the variations in local peaks and valleys, outperformed the other two frequencies parameters in classifying 40–60 voices of non-vacuity and deficient patients. The authors also proposed a reasonable choice of the utterance, such as the vowel 'a', to be pronounced by the patients for the discrimination purpose.

While the results presented in the paper are adequate, it will be interesting for future research to classify more groups of patients based on the parameters defined in this study and to explore additional parameters which may be suitable for distinguishing patients with different conditions.

The fractal dimension (FD) parameter was proposed to compute the speech characteristics related to production irregularities [34]. The difference in complexity in the test sentences, each pronounced three times by patients, was compared using the dynamic time warping (DTW) technique. Although the same sentence was read by the same patient three times, the distributions of FD values each time were not exactly the same to each other. Based on 9–16 voices for each class, the classification result between the deficient and normal patients exceeds 85%. The authors also claimed that the classification accuracy using FD values was more effective than the acoustic parameters suggested by Chiu et al. [33]. In these studies, we noticed that only a small set of data were used. However, much more data will usually be required to assure good performance in the fielded system. Since these studies were based on parametric approaches, it will be motivating to attempt nonparametric techniques or the integration of both.

4.3. Pulse analysis

The rationale for pulse analysis is based on the fact that blood travels at different pace at different organs, resulting in different shapes which allow characterization of the health condition of a particular organ. Pulse waves are usually measured at six points near the wrists of both of the patient's hands. The different points are correlated to different organs [35] as shown in Table 3. Subjective pulse diagnosis is very difficult to learn and many times accurate subjective pulse diagnosis can only be made by TCM practitioners after years of experiences. Therefore, a computational system to accurately diagnose pulse waves is highly desirable.

Before analyzing the pulse wave data using computational approaches, a measurement equipment needs to be used to record pulse waves. It comprises input of two parts, one is to measure the pulse waves using a condenser microphone with amplifiers and Bessel filters, allowing the transmission of the pulse waves into the computer via an A/D converter. The other part is to weigh the pressure when the condenser microphone is pressed against the patient's wrist. The condenser microphone uses the same principle as that of the phonocardiogram which is employed in the Western medicine to analyze the heart condition. Next, the input waves are analyzed using the graphical expression and the derivative method. The visual representation of the pulse waves facilitates inexperienced TCM practitioners to distinguish the disease and normal ones. A derivative method is used to normalize the pulse waves to allow clearer comparisons.

The advantage of pulse analysis includes a correlation study among different affected organs. While blood tests and ultrasound scanning enable detection of liver problems, there is no approach other than TCM pulse analysis which allows assessment of how liver problems affect other organs such as spleen, stomach, lung and gall bladder [36]. Their results for the 85 patients with possible liver problems indicated that the meridian theory and pulse analysis have physiological and pathological importance. Furthermore, the 24 patients with liver cirrhosis were found to possess splenomegaly and abdominal wall venous collateral, suggesting that the related organs such as spleen and spleen meridian are also affected in severe liver problems.

In a more recent work, a quantitative system for pulse diagnosis was built based on Bayesian networks to construct the mapping relationships between pulse wave parameters and pulse types [37]. The pulse is classified into seven types: floating, sinking, moderate, rapid, knotted, intermittent, feeble, replete, slippery, hesitant, and wiry, according to the following seven factors: depth, width, length, frequency, rhythm, quality, and strength. The probabilistic reasoning module based on Bayesian networks consists of discovering dependency relationship module, parameter learning and reasoning module. Their results showed that the system predicted the pulse types at 84% accuracy level. However, the pulse parameters utilized in the system might not be discriminated enough for diagnosing pulse types on small sample sizes, which have inherent deprivation of correspondence among pulse samples themselves.

4.4. TCM expert systems

TCM expert systems have great usefulness to medical practitioners and general publics. Expert systems allow generation of clinical alerts and accurate interpretations or diagnosis. The high erroneous rate of 80% for gastric ulcer treatment using TCM (World Health Organization, 2002) highlighted the need of reliable expert system for TCM. The design of such systems has been actively researched for years. Many TCM expert sytems embrace the four diagnostic methods: inspection, auscultation and olfaction, inquiring, and palpation. The choice of algorithm for knowledge acquisition or data mining, is a crucial consideration in the development of TCM expert system. The most important step is to create abstract rules from a large number of cases. A number of factors, such as problem definition, feasibility study, and user's demand investigation, must be considered in the design of an expert system. The techniques used for TCM expert system include Bayesian approach, weighted summation, and ontology. Some TCM expert systems and their approaches are summarized in Table 4 and discussed below. Many systems adopt hybrid approaches for better performance.

Motivated by the widely use of Bayesian methods in the medical system which posed some limitations such as low performance in solving the contradiction between the low frequency of a disease manifestation and the high specificity

Table 3 – Different pulse-measuring points on wrists and their corresponding organs		
	Right hand	Left hand
Cun	Middle chest, lung	Heart
Guan	Stomach, spleen	Gall, liver
Ci	Big intestine, left kidney	Small intestine, bladder, right kidney

Table 4 – TCM diagnosis systems and their approaches

System	Approaches	References
Electronic-Brain Medical Erudite (EBME)/Medical Expert System (MES)	Weighted summation, such as algebraic sum method	Xiao et al. [38], Yan et al. [39]
Chinese Medical Diagnostic System (CMDS)	Ontology	Huang and Chen [40]
A self-learning expert system	A hybrid Bayesian network learning algorithm, Nave–Bayes classifiers	Wang et al. [1]
Traditional Chinese Medicine Sizheng Integrated Recorder and aided syndrome Differentiator (TCM-SIRD)	Analysis packages of tongue and facial images, odor, speech and pulse	Zheng and Wu [41]

of the manifestation, the algebraic sum method, a kind of weighted summation method, was proposed in the development of the *Electronic-Brain Medical Erudite* (EBME) [38]. EBME has the Enormous Knowledge Base of Disease Diagnosis Criteria (EKBDDC) as its essence. EKBDDC depicts approximately 4000 items of diagnostic information for 1001 diagnostic entities [42]. Aimed at automating the knowledge engineering processes, avoiding repeated labor, and standardizing medical information processing, the description of each item of diagnostic information utilizes the most commonly used names and terms. Additionally, each diagnostic item has a unique one-to-one code, made possible by the creation of a diagnostic information base (DIB) in which about 5000 items of diagnostic information was preliminary prepared. A useful aspect of the system is a disease group diagnosis prior to a disease diagnosis. Each disease group comprises several, or a maximum of 30, diseases.

The improved diagnostic accuracy of using empirical method of weighted summation compared to that of using Bayesian algorithm [38], motivated the adoption of the approach for the construction of large-scale *Medical Expert System* (MES) [39]. Knowledge acquisition (KA) is usually the bottleneck of building knowledge-based system. A medical KA/management system is constructed using a three-tier distributed client/server architecture and the knowledge is stored in the system in three knowledge bases. An 8-digit numeric indexing system with a weight value is used to assist the knowledge representation in the knowledge bases and information retrieval on the Internet.

Chinese Medical Diagnostic System (CMDS) [40] contains an integrated medical ontology and was designed to be online. The role of the ontologies used in this system is to capture domain knowledge and provide a standardized understanding of a medical domain, hence facilitating acquisition, verification, and maintenance by both human and machines. The prototype of CMDS can diagnose about 50 types of diseases by using over 500 rules and 600 images for various diseases. CMDS also includes pictures of Chinese herbal medicine with appropriate explanation on their usages. CMDS can diagnose stomachache, vomiting, hiccups, diaphragmatitis, bellyache, diarrhea, dysentery, constipation, jaundice and tympanites.

Wang et al. [1] developed a *self-learning expert system* for diagnosis in TCM using a hybrid Bayesian network learning algorithm, Nave–Bayes classifiers with a novel score-based strategy for feature selection and a method for mining constrained association rules. The learned knowledge is provided in multiple forms comprising causal diagram, association and reasoning rules derived from classifiers.

Traditional Chinese Medicine Sizheng Integrated Recorder and aided syndrome Differentiator (TCM-SlRD) was developed to record the four diagnostic aspects of TCM [41]. The TCM-SIRD framework encompasses five analysis packages, they are Tongue Image Analysis Package (TIAP), Facial Image Analysis Package (FIAP), Odour Analysis Package (OAP), Speech Analysis Package (SAP), and Pulse-wave Analysis Package (PAP). The framework is divided into three parts: sensor modules, peripheral sensor circuits, and Sizheng Information Server. The sensor modules consist of a piezo-electricity transducer for measuring the pressure points, an odor sensor module and a microphone for collecting the voice of patients, and a CCD camera for capturing the images of patients' tongue and faces. The peripheral sensor circuits serve to transform the four types of diagnostics information obtained from the sensor modules to current or voltage signals. The Sizheng Information Server is used to configure sensor parameters and to merge the data. Although a clinical case of TCM-SIRD was mentioned, no quantification results were revealed on how accurate the TCM-SIRD was in TCM diagnosis.

5. Other biomedical mining systems for TCM

Other biomedical mining systems are not directly related to TCM formulations or diagnosis. Some systems aim to discover relationships among various TCM-related components. For example, the Individual Diagnosis and Treatment System (IDTS) aims to mine association rules between Chinese medicine symptoms and the indications of patients [43]. Wang et al. [44,45] combined cluster analysis and Bayesian networks to find causal relationships in TCM in a kidney study. Shim and Xu [46] proposed a data mining model for an oriental medicine, including TCM, through Bayesian Ying Yang (BYY) Binary Independent Factor Analysis (IFA). Other systems aim at finding the genes and proteins related to TCM symptom complexes from both TCM and modern biomedical literature by utilizing machine learning, text mining and web information techniques [47–49]. Yang et al. [50] proposed a Prior knowledge Support Vector Machine (P-SVM) to build an Information Management System of TCM Syndrome. A summary of the biomedical mining systems is given in Table 5.

In the Individual Diagnosis and Treatment System (IDTS), [43] generated results on the topic 'Relationship between Chi-

Table 5 – Other TCM biomedical mining systems		
Study	Approaches	References
Individual Diagnosis and Treatment System (IDTS)	Association rule mining	He et al. [43]
Causal discovery in TCM on a kidney study	Combination of hierarchical cluster approach and Bayesian networks	Wang et al. [44]
A data mining model for TCM	Bayesian Ying Yang (BYY) Binary Independent Factor Analysis (IFA) including harmony learning and model selection	Shim and Xu [46]
Discovery of genes and proteins related to TCM symptom complexes	Combination of machine learning, text mining and web information techniques	Wu et al. [47], Zhou et al. [48]
Information Management System of TCM Syndrome	Prior knowledge Support Vector Machine (P-SVM)	Yang et al. [50]

nese Medicine Symptom and the indication of patients' as association rules. Clinical data of 740 patients were gathered without restriction by Beijing Xi Yuan and another three hospitals. Then, data were cleaned up to ensure the uniqueness of the patients' ID, the drugs code standardization and medicinal preparation standardization. With association rule mining, it was observed that the number of rules generated decreased with higher confidence constraints and the support is low for most of the rules. It was suggested that the incomplete original data might contribute to the poor result.

Wang et al. [44] combined cluster analysis and Bayesian networks in TCM in a kidney study. The hierarchical cluster approach is used to assist causal discovery and Bayesian networks are used to explain relationships among variables in each cluster. A stepwise approach to network structural learning is proposed where a subnetwork over a larger cluster is constructed by combining several subnetworks over small clusters whenever these small clusters are grouped together.

Shim and Xu [46] proposed a data mining model for an oriental medicine, including TCM, through Bayesian Ying Yang (BYY) Binary Independent Factor Analysis (IFA) which was implemented in a three-layer structure that has two types of learning algorithms in two steps. In the first step, BYY harmony learning (i.e. to consider the joint distribution of x, y via two types of Bayesian decomposition of the joint density $q(z|y)q(y) = q(z, y)$, $p(z, y) = p(y|z)p(z)$) and model selection (i.e. to decide the number of k hidden independent factors) are proposed for Binary IFA. In the second step, a supervised learning algorithm is proposed which is capable of adaptive learning and can determine the number of independent factors and then separate them from the observed mixed data. They obtained an accuracy of 90.5% on diagnosing from 300 sample data.

Utilizing machine learning, text mining and web information techniques, the genes and proteins related to symptom complexes can be found from both TCM and modern biomedical literature, hence connecting both TCM and modern life sciences [47,48]. Symptom complex is a holistic clinical disease concept reflecting the dynamic, functional, and spatio-temporal disease status of human body. A symptom complex may occur in several different diseases whereas a disease may have several symptom complexes describing it. For example, Kidney YangXu refers to tens to hundreds diseases whereas diabetes can be described by several symptom complexes such as Kidney YangXu, YingYangLiangXu, and

QiYingLiangXu. There are some similarities between the characteristics of symptom complexes and those of genome and proteome as exemplified by polymorphism and dynamic associations of biological players. In fact, Kidney YangXu was reported to be associated with the expression of C1q-related factor (CRF) [49]. Data mining was performed on the following databases: Chinese medicine, formulation, chemical substance, and disease. These are then linked to biomedical literature database MEDLINE[4] to perform TCM data mining and knowledge discovery. Concurrently, information such as hospitals, drug products, and diseases information can be retrieved.

Zhou et al. [48] used disease principles as connecting points between symptom complex and genes/gene products. The information on evidence disease relationships such as the clinical diseases names, are retrieved using bootstrapping from TCM literature whereas the relationships between diseases and genes/gene products are retrieved from MEDLINE based on term co-occurrence. They acquired the gene nomenclature information from the HUGO Nomenclature Committee[5], which has 17,888 approved gene symbols. They hypothesized that the relevant genes belonging to the same symptom complex will have some biological interactions. In addition, since TCM symptom complex encompasses the concept of spatio-temporal functions, it is interesting to explore the interactions among genes and proteins in the context of symptom complexes, which will add a new perspective to the quest in system biology which aims to integrate biological data in order to understand how biological systems function. In their preliminary results, from 5061 disease evidence relationships, they managed to find out 110 terms of symptom complex, 634 disease terms, 13,277 gene–disease relationships which include 2958 gene terms. In a subsequent study, they discovered 790 terms of symptom complex, 2763 terms of disease terms, and 1100 pairs of disease and symptom complex. The results of seven relevant diseases of Kidney YangXu relating to genes in MEDLINE indicate that Kidney YangXu symptom complex is related to nervous system, immunology, and endocrine system.

Information Management System of TCM Syndrome was developed based on Prior knowledge Support Vector Machine

---

[4] http://www.nlc.nlm.ncbi.gov.
[5] http://www.gene.ucl.ac.uk/nomenclature/.

(P-SVM) [50], which was based on traditional SVM. P-SVM is used to perform classification of TCM syndrome literal information. P-SMO, a Sequential Minimal Optimization (SMO)-like algorithm [51], is used to train the model on the sample dataset of approximately 400,000 entries with confidence characteristic to produce the final classifier. Their results indicated that there is a significant improvement on classification accuracy, reaching 95% accuracy rate, with the use of strong expert knowledge.

6. Knowledge base grid for TCM

Grid is a computing model in which many networked computers are modelled as a virtual computer architecture to integrate services, share coordinated resources and solve problems across distributed, heterogeneous and multi-institutional virtual organizations with proper standards and security provisioning [52,53]. Knowledge Base Grid for TCM is useful for TCM knowledge discovery and exchange among TCM practitioners and researchers worldwide, instead of sole dependencies on databases provided by individual institutions. A domain-oriented TCM Knowledge Base Grid was developed on the semantic web to collect basic concepts, terms, rules, and cases of TCM [9,54]. They also suggested a Grid Ontology which attempts to enable Grid intelligence. The semantic web is based on Resource Description Framework (RDF) metadata techniques and XML data representation. The development of TCM Knowledge Base Grid was initially based on Open Grid Service Architecture (OGSA) [55], but they discovered that OGSA could not provide proper database resource integration, registration, and discovery. Thus, they introduce a framework comprising three components: Virtual Open Knowledge Base (VOKB), Knowledge Base Index (KBI), and Semantic Browser. VOKB comprises many distributed knowledge bases, which are built using many knowledge sources. The databases included are patient record databases, TCM medicine databases, traditional Chinese drug database, TCM literature databases, traditional Tibetan drug database, etc. KBI is the information depository of VOKBs and serves as a linker between VOKB and Semantic Browser. Semantic Browser supports TCM concepts browsing and TCM knowledge management. Depending on TCM ontology standard which can be a public reference, different VOKBs can communicate with each other. In the TCM Knowledge Base Grid, both the KBI and all the VOKBs are constructed based on the same TCM public ontology. In order to provide a unified web-accessible knowledge resources, the TCM ontology were rebuilt using Protégé 2000 [56], which was developed by Medical Informatics of Stanford University.

7. Future trends

Unprecedented growth in both the production of biomedical data and the amount of published literature on TCM research, highlights the importance of proper data mining. Although there are increasing efforts to make Chinese-language data sources as efficient and effective as possible, time-consuming and manual searchings are still predominant.

The holistic and systematic approach of TCM and the research of diseases, drugs and molecules of modern biology, will complement each other. Computational tools for knowledge discovery in these two fields will play a pivotal role in the improvements of human healths. Herein, we provided some important challenges crucial in TCM study.

While Western medicine currently employs pure, single compounds, TCM has long used multiple combinations of medicinal herbs to treat and relieve the symptoms of many different human diseases. The multi-component and synergistic nature of TCM is important in pharmaceutical development. It is crucial to identify and improve the efficacy of the active principals of herbs both singly and in combination—from active ingredients, active fractions, and active herbal formulations. There are thus complicated relationships among a large number of TCM ingredients, their efficacy, toxicity, indication, etc. Effective computational approaches need to be investigated to reveal the relationships among various TCM components.

Some preliminary studies have already been conducted to find out the relations between TCM and biomedicine. For example, disease principles could serve as connecting points between symptom complex and gene/gene products. Through text mining in both TCM literature and MEDLINE, it is possible to explore the interactions among genes and proteins in the context of symptom complex. This opens a new perspective in system biology which aims to integrate biological data in order to understand how biological system function. As one of the promising future directions, it is possible to explore gene regulatory networks or protein–protein interactions through the knowledge revealed from TCM symptom complex.

For TCM diagnosis, most proposed TCM expert systems rely on Bayesian networks (BNs) as it is natural to capture casual relationships in complicated features obtained in TCM diagnosis. Besides, prior knowledge can also be easily represented by BNs. However, it is well-known that BN topology learning becomes intractable when the number of features grows dramatically. It is therefore crucial to identify essential features relevant to TCM diagnosis through some effective computational methods.

Computational methods for TCM generate poorer results than those of Western medicine. The main reason is that information from ontologies are not well used. Although there have been several TCM approaches employing different ontologies, standardized ontology seems absolutely necessary. Moreover, TCM literature needs to be semantically annotated and actively linked to ontologies and knowledge in databases.

Computational approaches play bidirectional roles in bridging the gap between the TCM and Western medicine research. At present, the computational methods for TCM have mainly been developed in China, Hong Kong or Taiwan. Both computational and experimental efforts have been made to relate the TCM components and those of Western medicine, in order to demonstrate the effectiveness of TCM scientifically. At the same time, these relationships are crucial for defining standards for evaluating computational methods for TCM and identifying specific requirements and potential applications. Ultimately, integrated knowledge systems for querying and analysis will be delivered by computational methods. To achieve the aim of bridging their gap, computational tools have to be bilingual and user friendly so that they are acces-

sible by experts of both fields. Moreover, to deal with the exponentially growing data, application of Knowledge Base Grid for TCM will definitely be useful.

Finally, the development of standards for evaluating various computational tools for TCM is necessary as the construction of such evaluation benchmarks are potential to drive the significant developments of computational methods for TCM.

REFERENCES

[1] X. Wang, H. Qu, P. Liu, Y. Cheng, A self-learning expert system for diagnosis in traditional Chinese medicine, Expert Syst. Appl. 26 (2004) 557–566.

[2] R. Hoessel, S. Leclerc, J.A. Endicott, Indirubin, the active constituent of a Chinese antileukaemia medicine, inhibits cyclin-dependent kinases, Nat. Cell Biol. 1 (1999) 60–67.

[3] D.P. Reid, Chinese herbal medicine, CFW Publications Ltd., Hong Kong, 1997, p. 26.

[4] X. Qiao, T. Hou, H. Yu, X. Xu, Research and development of traditional Chinese medicine drugs, Acta Phys. Chim. Sin. 18 (2002) 394–398.

[5] X. Qiao, T. Hou, W. Zhang, S. Guo, X. Xu, A 3D structure database of components from Chinese traditional medicinal herbs, J. Chem. Inf. Comput. Sci. 42 (2002) 481–489.

[6] M. He, X. Yan, J. Zhou, G. Xie, Traditional Chinese medicine database and application on the web, J. Chem. Inf. Comput. Sci. 41 (2001) 273–277.

[7] X. Yan, Zhou J., Xie G., Traditional Chinese Medicines: Molecular Structures, Natural Sources and Applications, 1999.

[8] K.W. Fan, Online research databases and journals of chinese medicine, J. Altern. Complement. Med. 10 (2004) 1123–1128.

[9] J. Xu, H. Chen, Z. Wu, Knowledge base grid and its application in traditional Chinese medicine, IEEE Int. Conf. Syst. 3 (2003) 2477–2482.

[10] R. Grishman, B. Sundheim, Message understanding conference-6: a brief history, in: Proceedings of the 16th conference on Computational linguistics, vol. 1, 1996, pp. 466–471.

[11] Y. Tan, T. Yao, Q. Chen, J. Zhu, Applying conditional random fields to Chinese shallow parsing, in: Lecture Notes in Computer Science-Computational Linguistics and Intelligent Text Processing (CICLing), vol. 3406, 2005, pp. 167–176.

[12] Y. Guo, Chinese Question Answering with Full-Text Retrieval Re-Visited, University of Waterloo, 2004.

[13] T.-H. Tsai, S.-H. Wu, C.-W. Lee, C.-W. Shih, W.-L. Hsu, Mencius: a Chinese named entity recognizer using maximum entropy-based hybrid model, Comput. Linguist. Chin. Lang. Process. 9 (2004) 65–82.

[14] X. Zhou, Z. Wu, A. Yin, L. Wu, W. Fan, R. Zhang, Ontology development for unified traditional Chinese medical language system, Artif. Intell. Med. 32 (2004) 15–27.

[15] C. Li, C. Tang, J. Peng, J. Hu, L. Zeng, X. Yin, J. Jiang, J. Liu, TCMiner: a high performance data mining system for multi-dimensional data analysis of traditional chinese medicine prescriptions, Lect. Notes Comput. Sci. 3289 (2004) 246–257.

[16] X. Zhou, B. Liu, Z. Wu, Text mining for clinical Chinese herbal medical knowledge discovery, Lect. Notes Comput. Sci. 3735 (2005) 396–398.

[17] C. Cao, H. Wang, Y. Sui, Knowledge modeling and acquisiton of traditional Chinese herbal drugs and formulae from text, Artif. Intell. Med. 32 (2004) 3–13.

[18] F. Zhang, D. Li, H. Yang, D. Hu, Research of Chinese traditional medicine compound innovation system, Comput. Data Eng. 32 (2004) 9–12.

[19] R. Agrawal, R. Srikant, Fast algorithms for mining association rules in large databases, in: Proceedings of the 20th International Conference on Very Large Data Bases, 1994, pp. 487–499.

[20] J. Han, J. Pei, Y. Yin, Mining frequent patterns without candidate generation, in: ACM-SIGMOD International Conference on Management of Data, vol. 29, 2000, pp. 1–12.

[21] G.G. Zhang, W. Lee, B. Bausell, L. Lao, B. Handwerger, B. Berman, Variability in the traditional chinese medicine (TCM) diagnoses and herbal prescriptions provided by three TCM practitioners for 40 patients with rheumatoid arthritis, J. Altern. Complement. Med. 11 (2005) 415–421.

[22] F.L. Zhu, C. Gu, Z.F. Zhang, C.L. Zhou, J.T. Xu, X.G. Liu, W.S. Li, H.Q. Zhu, Evaluation of therapeutic effects of Chinese materia medica by tongue image analysis software 1.0 based on tongue colors, Zhong Xi Yi Jie He Xue Bao 4 (2006) 152–155.

[23] B. Pham, Y. Cai, Visualization techniques for tongue analysis in traditional Chinese medicine, in: Proceedings The Medical Imaging 2004: Visualization, Image Guided Procedures and Display, San Diego, California, 2004.

[24] C.-C. Chiu, A novel approach based on computerized image analysis for traditional Chinese medical diagnosis of the tongue, Comput. Methods Programs Biomed. 61 (2000) 77–89.

[25] T. Watsuji, S. Arita, S. Shinohara, T. Kitade, Medical application of fuzzy theory to the diagnostic system of tongue inspection in traditional Chinese medicine, in: IEEE International Fuzzy Systems Conference Proceedings, vol. 1, 1999, pp. 145–148.

[26] J.H. Jang, J.E. Kim, K.M. Park, S.O. Park, Y.S. Chang, B.Y. Kim, Development of the digital tongue inspection system with image analysis, in: Joint Engineering in Medicine and Biology/Biomedical Engineering Society Conference, vol. 2, 2002, pp. 1033–1034.

[27] B. Pang, D. Zhang, N. Li, K. Wang, Computerized tongue diagnosis based on Bayesian networks, IEEE Trans. Biomed. Eng. (2004) 1803–1810.

[28] H.Z. Zhang, K.Q. Wang, D. Zhang, B. Pang, B. Huang, Computer aided tongue diagnosis system, IEEE Eng. Med. Biol. (2005) 6754–6757.

[29] J. Wu, Y. Zhang, J. Bai, Tongue area extraction in tongue diagnosis of traditional Chinese medicine, in: IEEE International Conference of the Engineering in Medicine and Biology Society, 2005, pp. 4955–4957.

[30] Y. Deng, B.S. Manjunath, Unsupervised segmentation of color-texture regions in images and video, IEEE Trans. Pattern Anal. Mach. Intell. 23 (2001) 800–810.

[31] Y. Wang, Y. Zhou, J. Yang, Y. Wang, JSEG based color separation of tongue image in traditional Chinese medicine, Lect. Notes Comput. Sci. Prog. Pattern Recognit. Image Anal. Appl. 3287 (2004) 503–508.

[32] B. Pang, D. Zhang, K. Wang, Tongue image analysis for appendicitis diagnosis, Inf. Sci. 175 (2005) 160–176.

[33] C.-C. Chiu, H.-H. Chang, C.-H. Yang, Objective auscultation for traditional Chinese medical diagnosis using novel acoustic parameters, Comput. Methods Programs Biomed 62 (2000) 99–107.

[34] C.-C. Chiu, M.-T. Yang, C.-S. Lin, Using fractal dimension analysis on objective auscultation of traditional Chinese medical diagnosis, J. Med. Biol. Eng. 22 (2002) 219–225.

[35] W.C. Tang, R. Li, Research on the Cun-Guan-Chi pulse detecting system, Zhongguo Yi Liao Qi Xie Za Zhi 29 (2005) 164–166.

[36] W.A. Lu, Y.Y. Lin Wang, W.K. Wang, Pulse analysis of patients with severe liver problems, IEEE Eng. Med. Biol. 18 (1999) 73–75.

[37] H. Wang, Y. Cheng, A quantitative system for pulse diagnosis in traditional Chinese medicine, in: IEEE International Conference of the Engineering in Medicine and Biology Society, 2005, pp. 5676–5679.

[38] S. Xiao, C. Peng, Z. Wang, F. Wang, Using the algebraic sum method in medical expert systems, IEEE Eng. Med. Biol. (1996) 80–82.

[39] H. Yan, Y. Jiang, J. Zheng, B. Fu, S. Xiao, C. Peng, The internet-based knowledge acquisition and management method to construct large-scale distributed, Comput. Methods Programs Biomed. 74 (2004) 1–10.

[40] M. Huang, M. Chen, Integrated design of the intelligent web-based Chinese medical diagnostic system (CMDS)—systematic development for digestive health, Expert Syst. Appl. 32 (2007) 658–673.

[41] N. Zheng, Z. Wu, TCM-SIRD: an integrated aided system for traditional Chinese medicine sizheng, IEEE Int. Conf. Syst. Man Cybern. 4 (2004) 3864–3868.

[42] S. Xiao, F. Wang, Building knowledge-based systems with an assembling technique, IEEE Eng. Med. Biol. 16 (1997) 81–83.

[43] X. He, W. Huang, M. Lu, W. Xue, Y. Lu, Research and application of data mining in individual diagnosis and treatment based on Chinese traditional medicine, in: IEEE International Conference on Granular Computing, 2006, pp. 417–419.

[44] M. Wang, Z. Geng, M. Wang, F. Chen, W. Ding, M. Liu, Combination of network construction and cluster analysis and its application to traditional Chinese medicine, Lect. Notes Comput. Sci. 3973 (2006) 777–785.

[45] K. Deng, D. Liu, S. Gao, Z. Geng, Structural learning of graphical models and its applications to traditional Chinese medicine, in: L.P. Wang, Y.C. Jin (Eds.), Proceedings of Fuzzy Systems and Knowledge Discovery 2005, LNCS 3614, Springer-Verlag, Berlin, 2005, pp. 362–367.

[46] J. Shim, L. Xu, Medical data mining model for oriental medicine via BYY binary independent factor analysis, in: Proceedings of the 2003 International Symposium on Circuits and Systems, vol. 5, 2003, pp. 717–720.

[47] Z. Wu, X. Zhou, J. Chen, B. Liu, Text mining for finding functional community of related genes using TCM knowledge, in: Proceedings of the 8th European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD), 2004, pp. 459–470.

[48] X. Zhou, Z. Wu, B. Liu, Knowledge discovery in biomedical literature: survey and prospect, Complex Syst. Complex. Sci. 1 (2004) 45–55.

[49] Z. Shen, The continuation of kidney study, Shanghai Scientific & Technical Publishers, Shanghai, 1990.

[50] X. Yang, Z. Liang, G. Zhang, Y. Luo, J. Yin, A classification algorithm for TCM syndromes based on P-SVM, Int. Conf. Mach. Learn. Cybern., (2005) 3692–3697.

[51] J.C. Platt, Fast training of support vector machines using sequential minimal optimization, in: Advances in Kernel Methods: Support Vector Learning, MIT Press, 1999.

[52] M. Klein, XML, RDF, and relatives, IEEE Intell. Syst. 16 (2001) 26–28.

[53] I. Foster, C. Kesselman, S. Tuecke, The anatomy of the grid: enabling scalable virtual organizations, Int. J. High Perform. Comput. Appl. 15 (2001) 200–222.

[54] J. Xu, Z. Wu, A survey of knowledge base grid for traditional Chinese medicine, Enterprise Inf. Syst. V (2005) 316–322.

[55] H. Chen, Z. Wu, C. Huang, Open grid services of traditional Chinese medicine, IEEE Int. Conf. Syst. Man Cybern. 5 (2003) 4546–4551.

[56] N.F. Noy, M.A. Musen, Creting semantic web contents with Protégé 2000, IEEE Intell. Syst. 16 (2001) 60–71.