Research Article

Application of Multilabel Learning Using the Relevant Feature for Each Label in Chronic Gastritis Syndrome Diagnosis

Guo-Ping Liu, Jian-Jun Yan, Yi-Qin Wang, Jing-Jing Fu, Zhao-Xia Xu, Rui Guo, and Peng Qian

1 Laboratory of Information Access and Synthesis of TCM Four Diagnosis, Basic Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China
2 Center for Mechatronics Engineering, East China University of Science and Technology, Shanghai 200237, China

Correspondence should be addressed to Jian-Jun Yan, jjyan@ecust.edu.cn and Yi-Qin Wang, wangyiqin2380@sina.com

Received 13 January 2012; Accepted 22 March 2012

Copyright © 2012 Guo-Ping Liu et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Background. In Traditional Chinese Medicine (TCM), most of the algorithms are used to solve problems of syndrome diagnosis that only focus on one syndrome, that is, single label learning. However, in clinical practice, patients may simultaneously have more than one syndrome, which has its own symptoms (signs).

Methods. We employed a multilabel learning using the relevant feature for each label (REAL) algorithm to construct a syndrome diagnostic model for chronic gastritis (CG) in TCM. REAL combines feature selection methods to select the significant symptoms (signs) of CG. The method was tested on 919 patients using the standard scale.

Results. The highest prediction accuracy was achieved when 20 features were selected. The features selected with the information gain were more consistent with the TCM theory. The lowest average accuracy was 54% using multi-label neural networks (BP-MLL), whereas the highest was 82% using REAL for constructing the diagnostic model. For coverage, hamming loss, and ranking loss, the values obtained using the REAL algorithm were the lowest at 0.160, 0.142, and 0.177, respectively.

Conclusion. REAL extracts the relevant symptoms (signs) for each syndrome and improves its recognition accuracy. Moreover, the studies will provide a reference for constructing syndrome diagnostic models and guide clinical practice.

1. Introduction

Although Traditional Chinese Medicine (TCM) and Western Medicine diagnose cases in clinical applications, their theoretical systems are totally different. The hard targets such as laboratory and imaging tests are very important for diagnosing diseases in Western Medicine whereas soft targets are much more important in the clinical diagnosis of TCM.

The so-called soft targets [1] mainly refer to targets that cannot be accurately measured and with poor repeatability. These targets are subjective and are collected through clinical observation of a doctor or the patient’s self-report, which cannot be accurately measured using instruments or directly through other means.

Therefore, collecting information in the diagnosis of TCM is difficult because it cannot be measured accurately, with poor measurement repeatability, and are easily influenced by the study sample and environmental factors.

Given that the soft targets of TCM are subjective, fuzzy, and multidimensional, TCM has been considered as a mystical experience in the scientific world and has not been identified in a wide range.

In recent years, the standardization and objectification of TCM diagnosis have gradually become a research hotspot with the development of mathematical statistics, data mining, and pattern recognition technology.

The studies are revealing the rules between syndromes and the information of four diagnosis: inspection, auscultation and olfaction, inquiring, and palpation, and seeking information of four diagnosis for differential diagnosis or extracting classification rules for syndrome differentiation. The research can provide a reference for clinical syndrome
differentiation and reduce the subjectivity and ambiguity of diagnosis.

Some researchers have applied the structural equation model in studying chronic atrophic gastritis. The results show that chronic atrophic gastritis resulting from the most common syndromes correspond with diagnostic targets, which is agreement with the clinical practice of TCM [2]. An improved conjugate gradient learning algorithm is used to create the BP model with three layers for diabetes and nephropathy. The results show its advantages in predicting diabetes and nephropathy [3]. An entropy-based partition method for complex systems is applied to establish endothelial dysfunction diagnostic criteria for Yin deficiency syndrome. Moreover, the experimental results are highly consistent with the findings of clinical diagnosis [4]. Multilabel learning [5] combined with the frequency method is presented to select the symptoms that greatly contribute to coronary heart disease. The results show the improvement in the diagnosis of coronary heart disease. Su et al. [6] employed the correlation coefficient, similarity D, the angle cosine, and spectral similarity to study the correlation between the symptoms (signs) and the five syndromes of liver cirrhosis. The research can provide a basis for differentiating patients with nonspecific clinical manifestations.

Our research group focuses on the standardization and objectification of syndrome in TCM. We applied latent structure models [7] to study the association between symptoms of the spleen system. According to a social network theory, we used the associated density method [8, 9] to analyze the correlation between syndrome-syndrome of coronary heart disease and symptom-syndrome of chronic gastritis.

In the studies on syndrome standardization and objectification mentioned above, most of the algorithms were used to solve problems in diagnosing patients with disjoint syndromes, which belong to single-label learning. However, in clinical practice, strong relevance may be observed among different syndromes. Traditional single-label data-mining techniques, which could only build one model at a time, ignore the fact that one patient may be associated with more than one syndrome. In this study, a novel multilabel learning (MLL) technique is explored to solve this problem. Our group [5] applied a multilabel learning algorithm (ML-kNN) to construct a syndrome model for diagnosing CHD in TCM. The ML-kNN produces better results than the ranking Support Vector Machine (Rank-SVM), BPMLL, and kNN based on three criteria, namely, average precision, coverage, and ranking loss.

Compared with traditional learning methods, multilabel learning more effectively identifies syndrome information in TCM and solves problems such as single samples with several syndromes. However, the relationship between features and class labels is not concerned in multilabel learning.

Chronic gastritis (CG) is a common disease and is classified under spleen and stomach diseases in TCM. According to preliminary studies, we applied the feature selection methods to select significant symptoms (signs) associated with each syndrome in CG. In addition, we further applied multilabel learning algorithm to construct the syndrome models of inquiry diagnosis for CG in TCM to provide a reference for the syndrome standardization and objectification of CG.

In this paper, the first section includes the introduction of the research progress in the field of TCM diagnosis, the purpose of the study, and its significance. In Section 2, we introduce the data-collecting methods, which include a variety of feature selection methods and a multilabel learning method designated as REAL. Standardizing a scale of inquiry information is discussed in the results section. The optimal symptom set is obtained for each syndrome using feature selection. The results of diagnostic models constructed are discussed based on the REAL method. Then the results of REAL are compared with other multilabel learning algorithms. Section 3, the results of feature selection and diagnostic models are analyzed based on TCM theory. The last section concludes and indicates several issues for future studies.

2. Material and Methods

2.1. Research Subjects. Chronic gastritis (CG) samples were collected from a clinic, in-patient department, and gastroscopy room of the digestive system department of the Longhua Hospital and the Shuguang Hospital of Shanghai University of Traditional Chinese Medicine, the Xinhua Hospital, the Putuo District Central Hospital, and the Shanghai Hospital of Traditional Chinese Medicine. This work was approved by the Shanghai Society of Medical Ethics. All patients signed an informed consent form. A total of 919 valid subjects were enrolled after excluding cases with TCM inquiry diagnosis scales that lacked information or cannot be diagnosed with CG. Among the 919 patients, 354 were male (38.5%, with an average age of 44.61 yr ± 14.54 yr) and 565 were female (61.5%, with an average age of 48.70 yr ± 12.74 yr).

2.1.1. Inclusion Criteria. Including criteria were

(1) patients who meet the diagnostic standards for CG and TCM syndromes, and
(2) patients who were informed and have agreed to join this investigation.

2.1.2. Diagnostic Standards.

Western Diagnostic Standards. The Consensus of National Seminar on CG held by the Chinese Medical Association Digestive Diseases Branch in 2006 [10] was referenced to diagnose whether a patient has CG based on gastroscopy results, pathologic results, and clinical performance.

2.1.3. Exclusion Criteria. Excluding criteria were:

1. mentally ill patients and patients with other severe systemic diseases,
2. patients who have difficulty in describing their conditions, and
3. patients who are not informed or refuse to cooperate.

2.2. Method for Establishing TCM Inquiry Diagnosis Scales. The research group was composed of Shanghai senior clinical experts on the digestive system, clinical doctors, and researchers. The final TCM inquiry diagnosis scales were drafted based on past experience in the production of scales [13], a wide range of literature about TCM spleen and stomach diseases, related documents in core magazines, and journals for over 15 years and reports about the frequency of symptoms associated with syndromes in CG diseases in TCM. The scales were also amended and fixed by two rounds of expert consultation and statistical tests. The scales include eight dimensions such as cold or heat, sweat, head, chest and abdomen, urine and stool, diet and taste, sleep, mood, woman aspects and contents of disease history, inspection, and palpation. More than 113 variables were ultimately included in these scales.

2.3. Investigation Methods. The clear definitions of symptoms, the specific methods, and the order of inquiry diagnosis are given in the scales. All samplers must have undergone unified training. The group members assemble regularly and discuss the information of typical patients to ensure the consistency of the collected data.

2.4. Diagnosis Methods. Three senior chief doctors with plenty of experience in clinical practices were invited for inquiry diagnosis of the cases in terms of the CG diagnostic standards made by our research group. If two of them have the same diagnosis results, the case was included. Otherwise, the case was not adopted until at least two of them came to the same conclusion.

2.5. Data Input and Process Methods. We have the following methods:

1. Build a database with Epidata software.
2. Input data two times independently.
3. The Epidata software compares the two data sets and checks out mistakes.
4. Check the investigation form logically in case of filling errors.

2.6. Feature Extraction Methods. To obtain the proper set of symptoms for each syndrome, we employed four feature selection methods, namely, mutual information (MI) [14], information gain (IG) [15, 16], conditional mutual information method (CMIM) [17], and minimum redundancy maximum relevance (mRMR) [18], to investigate the relationship between the symptoms and the six common syndromes (patterns), such as the accumulation of damp heat in the spleen-stomach, dampness obstructing the spleen-stomach, spleen-stomach qi deficiency, spleen-stomach cold deficiency, liver stagnation, and stagnated heat in the liver-stomach.

2.7. Multilabel Learning Methods. Many real-world problems involving ambiguous objects lose useful information when analyzed using the traditional single-label algorithm. Thus, it will be harmful to the learning performance. To minimize this information loss, multilabel learning was proposed. Most traditional multilabel classification approaches to learning methods in vector spaces are used based on the assumption that the instances should have the same set of features in the input space for each label. However, for specific labels, not all the features have strong correlations. ML-kNN is the lazy multilabel learning algorithm based on k-nearest neighbor techniques (kNN) [19]. Similar to the kNN algorithm, it finds the k nearest neighbors for each test instance; however, in ML-kNN, the label of each test instance is estimated directly using the k nearest neighbors in instance. We applied a new algorithm called REAL to fit the characteristics for inquiry diagnosis in TCM based on ML-kNN. The REAL algorithm extracts the best feature subset correlated with a certain label as its input space and then calculates the posterior probability combined with the ML-kNN algorithm. The REAL algorithm is shown in Algorithm 1.

2.8. Experimental Design and Evaluation. Different characteristics were selected using the REAL algorithm. We selected 112, 100, 70, 60, 50, 40, 30, 20, 10, and 5 symptoms (signs), which correlated with each syndrome to build a syndrome model to study the influence of the different symptoms (signs) on the diagnostic model.

Considering each example could simultaneously be associated with multiple labels, performance evaluation in multilabel learning is different from single-label learning. The following five multilabel evaluation parameters presented in [20] are used in this paper.

Average Precision. It evaluates the average fraction of labels ranked above a particular label \( y \in Y \), which actually are in \( Y \). The performance is perfect when \( \text{avgprec}_k(f) = 1 \); the
bigger the value of avgprec\(_{s}\)(f), the better the performance one has.

\[
\text{avgprec}\_{s}(f) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{|Y_i|} \times \left| \left\{ y' \mid \text{rank}_f(x_i, y') \leq \text{rank}_f(x_i, y), y' \in Y_i \right\} \right| \times \frac{\text{rank}_f(x_i, y)}{\text{rank}_f(x_i, y)}.
\]

Coverage. It evaluates how far on the average we need to go down the list of labels to cover all the proper labels of the instance. The smaller the value of coverage\(_{s}\)(f), the better the performance one has.

\[
\text{coverage}\_{s}(f) = \frac{1}{p} \sum_{i=1}^{p} \max_{y \in Y_i} \text{rank}_f(x_i, y) - 1,
\]

Ranking Loss. It evaluates the average fraction of label pairs that are reversely ordered for the instance. The performance is perfect when rloss\(_{ss}\)(f) = 0; the smaller the value of rloss\(_{ss}\)(f), the better the performance,

\[
\text{rloss}_{ss}(f) = \frac{1}{p} \sum_{i=1}^{p} \frac{1}{|Y_i|} \times \left| \left\{ (y_1, y_2) \mid f(x_i, y_1) \leq f(x_i, y_2), (y_1, y_2) \in Y_i \times Y_i \right\} \right|,
\]

where \(Y\) denotes the complementary set of \(Y\) in \(y \cdot y = \{1, 2, \ldots, Q\}\) be the finite set of labels.

Hamming Loss. It evaluates how many times instance-label pairs are misclassified; that is, a label not belonging to the instance is predicted, or a label belonging to the instance is not predicted one has,

\[
\text{hloss}_{r}(f) = \frac{1}{m} \sum_{i=1}^{m} \frac{1}{n} \times |f(x_i) \Delta Y_i|,
\]

where \(\Delta\) denotes the symmetric difference between two sets.

**Algorithm 1:** REAL algorithm.

Step 1: The best feature subset is extracted for each label based on feature selection.
Step 2: Search for the k nearest neighbors based on the distance between two training instances for each labels, which is calculated in the corresponding feature subspace instead of the whole feature space.
Step 3: Calculate the posterior probability with the k nearest neighbors and further the confidential threshold value.
Step 4: Estimate the posterior probabilities of test instances.

One-error. It evaluates how many times the top-ranked label is not in the set of proper labels of the instance. The performance is perfect when one-error\(_r\)(f) = 0 we have.

\[
\text{one-error}_r(f) = \frac{1}{m} \sum_{i=1}^{m} \left[ \arg \max_{y \in Y} f(x_i, y) \right] \not\in Y_i.
\]

For any predicted \(\pi_1, \pi_2\) equals 1 if \(\pi_1\) holds and 0 if otherwise. Note that, for single-label classification problems, a one-error is identical to an ordinary classification error.

3. Results and Discussion

3.1. The Results of the Finest Symptoms (Signs) Subsets

3.1.1. The Results of Finest Subsets of Specific Symptoms (Signs). In the research of multilabel classification of syndrome diagnosis for CG, feature selection methods such as mutual information, IG, CMIM, and MRMR were combined with multilabel learning. The prediction accuracy was highest when 20 features were selected for classification. Based on the results, the features selected by IG are more suitable for TCM theory than those using other algorithms.

(i) 12 specific symptoms (signs), including yellow tongue coating and greasy tongue coating, were extracted for the pattern of damp heat accumulating in the spleen-stomach.

(ii) 12 specific symptoms (signs), including white and greasy tongue coating, were extracted for the pattern of dampness obstructing the spleen-stomach.

(iii) 11 specific symptoms (signs), including fatigue and tongue with teeth marks, were extracted for the pattern of spleen-stomach qi deficiency.

(iv) 8 specific symptoms (signs), including cold limbs and preference for warm temperature, were extracted for the pattern of spleen-stomach cold deficiency.

(v) 9 specific symptoms (signs), including white aggravating after anxiety or anger, distending pain in the chest, and hypochondriac area, were extracted for the pattern of pattern of liver stagnation.

(vi) 12 specific symptoms (signs), including burning pain and red tongue, were extracted for the pattern of stagnated heat in liver-stomach.

The detailed information about symptoms (signs) is displayed in Table 1.
### Table 1: The finest subsets of specific symptoms (signs).

<table>
<thead>
<tr>
<th>Symptoms (signs)</th>
<th>Syndromes (patterns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damp-heat accumulating in the spleen-stomach</td>
<td>Dampness obstructing the spleen-stomach</td>
</tr>
<tr>
<td>Yellow tongue coating</td>
<td>Fatigue</td>
</tr>
<tr>
<td>Greasy tongue coating</td>
<td>Spleen-stomach qi deficiency</td>
</tr>
<tr>
<td>Spleen-stomach qi deficiency</td>
<td>Stagnated heat in liver-stomach</td>
</tr>
<tr>
<td>Dampness obstructing the spleen-stomach</td>
<td>Preference for warm</td>
</tr>
<tr>
<td>Spleen-stomach deficiency cold</td>
<td>Distending pain in the chest and hypochondriac area</td>
</tr>
<tr>
<td>Liver stagnation</td>
<td>Burning pain</td>
</tr>
<tr>
<td>Stagnated heat in liver-stomach</td>
<td>Preference for cold</td>
</tr>
<tr>
<td>Yellow tongue coating</td>
<td>Pale-white tongue</td>
</tr>
<tr>
<td>Thick tongue coating</td>
<td>Tongue with teethmarks</td>
</tr>
<tr>
<td>White tongue coating</td>
<td>Fat tongue</td>
</tr>
<tr>
<td>Tongue with teethmarks</td>
<td>Fat tongue</td>
</tr>
<tr>
<td>White tongue coating</td>
<td>Red tongue</td>
</tr>
<tr>
<td>White tongue coating</td>
<td>Preference for warm</td>
</tr>
<tr>
<td>Tongue with teethmarks</td>
<td>Preference for cold</td>
</tr>
<tr>
<td>Fat tongue</td>
<td>Light tongue</td>
</tr>
<tr>
<td>Red tongue</td>
<td>Pale-white tongue</td>
</tr>
<tr>
<td>Tongue with teethmarks</td>
<td>Pale-white tongue</td>
</tr>
<tr>
<td>Fat tongue</td>
<td>Red lips</td>
</tr>
<tr>
<td>White tongue coating</td>
<td>Mixed yellow and white tongue coating</td>
</tr>
<tr>
<td>Tongue with teethmarks</td>
<td>Slippery pulse</td>
</tr>
<tr>
<td>Thirsty</td>
<td>Cold limbs</td>
</tr>
<tr>
<td>Bitter taste in the mouth</td>
<td>Bluish or purple tongue</td>
</tr>
<tr>
<td>Preference for cold</td>
<td>Slippiness pulse</td>
</tr>
</tbody>
</table>

#### 3.1.2. The Results of Finest Subsets of Negative Symptoms (Signs). They are as follows

(i) 8 negative symptoms (signs) including white and thin tongue coating were extracted for the pattern of damp-heat accumulating in the spleen-stomach.

(ii) 12 negative symptoms (signs) including red and dark red tongue were extracted for the pattern of dampness obstructing the spleen-stomach.

(iii) 11 negative symptoms (signs) including thick tongue coating, mixed yellow, and white tongue coating were extracted for the pattern of spleen-stomach qi deficiency.

(iv) 8 negative symptoms (signs) including red lips and good appetite, but easily gets hungry, were extracted for the pattern of spleen-stomach cold deficiency.

(v) 9 negative symptoms (signs), including fatigue and pain when exposed to cold, were extracted for the pattern of liver stagnation.

(vi) 12 negative symptoms (signs) including tongue with teeth marks and fat tongue were extracted for the pattern of stagnated heat in liver-stomach.

The detailed information about negative symptoms (signs) is displayed in Table 2.

#### 3.2. The Results of Syndrome Classification Using Multilabel Learning Methods

##### 3.2.1. Comparison of Average Accuracy with Different Number of Features

Using the REAL algorithm, we selected 112, 100, 70, 60, 50, 40, 30, 20, 10, and 5 symptoms (signs), which correlated with each syndrome to build a syndrome classification model to study the influence of the different symptoms (signs) on the diagnostic model.

The abscissa represents the number of the selected features, and the vertical axis represents their prediction accuracy in Figure 1.

As shown in Figure 1, the average accuracy changes with the number of symptoms (signs). When the number...
Table 2: The finest subsets of negative symptoms (signs).

<table>
<thead>
<tr>
<th>Symptoms (Signs)</th>
<th>Syndromes (Patterns)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Damp-heat accumulating in the spleen-stomach</td>
<td>Spleen-stomach qi deficiency</td>
</tr>
<tr>
<td>1</td>
<td>White tongue coating</td>
</tr>
<tr>
<td>2</td>
<td>Thin tongue coating</td>
</tr>
<tr>
<td>3</td>
<td>Fat tongue</td>
</tr>
<tr>
<td>4</td>
<td>Tongue with teethmarks</td>
</tr>
<tr>
<td>5</td>
<td>Whitish Lips</td>
</tr>
<tr>
<td>6</td>
<td>Whitish complexion</td>
</tr>
<tr>
<td>7</td>
<td>Whitish tongue</td>
</tr>
<tr>
<td>8</td>
<td>Dark-purple lips</td>
</tr>
<tr>
<td>9</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: The average accuracy rate with different number of symptoms (signs) by using REAL methods.

As indicated in Table 3, the highest was 82%, obtained by REAL, whereas the lowest average precision was 54%, obtained using BP-MLL. For the indicators coverage, hamming loss, and ranking loss, the values obtained using the REAL algorithm were lowest at 0.160, 0.142, and 0.177, respectively. In summary, the results obtained using the REAL algorithm were the most accurate.

3.2.2. Comparison of Performance of Different Multilabel Learning Algorithms. We selected 20 symptoms (signs) to build the models and compared the five evaluation parameters obtained using ML-kNN, Ensembles of Classifier Chains (ECCs), BSVM, BP-MLL, Rank-SVM, and REAL algorithms. The result is shown in Table 3.

3.2.3. The Comparison of Accuracy Rates of Various Syndromes Using Different Multilabel Methods (the 20 Features Are Selected in REAL Method). The results of the REAL method were compared with the other multilabel learning methods, namely, BP-MLL, Rank-SVM, ECC, BSVM, and ML-kNN. The recognition accuracies of the six common syndromes of CG are shown in Table 4.

As shown in Table 4, for the pattern of damp heat accumulation in the spleen-stomach, the REAL algorithm achieved the highest accuracy rate, followed by ECC, BSVM, Rank-SVM, ML-kNN, and BP-MLL. For the pattern of dampness obstructing the spleen-stomach, the REAL
algorithm also had the highest accuracy rate, followed by BSVM, ECC, Rank-SVM, ML-kNN, and BP-MLL.

For the pattern of spleen-stomach qi deficiency, the accuracy rate obtained from ECC was the highest, followed by BSVM, REAL, ML-kNN, Rank-SVM, and BP-MLL. For the pattern of spleen-stomach cold deficiency, REAL, ML-kNN and BP-MLL had the highest accuracy rate at 96.6%, followed by ECC, BSVM, and Rank-SVM. For the pattern of liver stagnation, the REAL algorithm achieved the highest accuracy rate, followed by BP-MLL, ML-kNN, BSVM, ECC, and Rank-SVM. For the pattern of stagnated heat in the liver-stomach, BP-MLL and REAL algorithm achieved the highest accuracy rate, followed by ML-kNN, ECC, BSVM, and Rank-SVM.

From the results, the comprehensive performance of REAL method was the best, with the accuracy rates in the six syndromes, except for the pattern of spleen-stomach qi deficiency.

### 3.3. Discussion

A syndrome is a unique TCM concept. It is an abstractive conception of a variety of symptoms and signs. It is a pathological summarization of a certain stage of a disease, and it covers disease location, etiology, and the struggle between the body’s resistance and pathogenic factors. Different syndromes have different clinical manifestations.

Symptoms, which are the external manifestations of a disease and a syndrome, refer to subjective abnormalities and the abnormal signs of patients elicited by doctors using the four diagnostic methods.

The etiology, location, nature, the struggle between the body’s resistance and pathogenic factors, and the condition at a certain stage of the disease process are highly summarized using syndrome differentiation. Syndrome differentiation involves three steps: (a) determining symptoms and signs through inspection, auscultation, inquiry, and palpation; (b) making an overall analysis of the information; (c) making a diagnostic conclusion. All these steps are based on TCM theory.

Figure 2 shows the TCM diagnosis of the network structure diagram. Network structure can be compared to a tree, where the root node is composed of a number of leaf nodes. $X_1, X_2, \ldots, X_7$ leaf nodes are directly observed, and we call them manifest variables, which denote the symptoms like latent variables $Z_1, Z_2, Z_3, Z_4$ are the root nodes that are indirectly measured through their manifestations, and we call them latent variables, which represent the syndromes of chronic gastritis in TCM. The syndrome can be observed alone or with others, such as $Z_1$ and $Z_2$, or $Z_2$ and $Z_3$, which may appear together.

$D$ denotes the disease. In this study, it represents chronic gastritis, which is a disease defined in Western Medicine. Chinese medical diagnosis of chronic gastritis may contain syndromes like latent variables $Z_1, Z_2, Z_3 \ldots$ and so on.

### 3.3.1. The Finest Symptoms (Signs) Feature Subsets for Each Syndrome

Feature selection is a hot topic in the field of machine learning. It studies how to select the most effective feature subset from a set of original feature sets to reduce
spleen-stomach qi deficiency syndrome usually exists with
is mainly composed of several syndromes. For example,
dromes may exist. The syndrome complex of one patient
associated with several syndromes.

In multilabel data, there is a relationship among labels. However, this relationship may be bound to be ignored
inherently by using the single-label learning. For this reason,
multilabel learning algorithms are developed to facilitate the
correlation of the labels.

Compared with other traditional multilabel learning
methods, the REAL algorithm found the relevant symptom subset of each syndrome with feature selection. Moreover,
the REAL algorithm identified the syndrome information of
TCM more effectively and accurately.

In addition, the REAL algorithm assisted in extracting
the corresponding specificity and negative symptoms (signs)
through feature selection. Extracted features are not only
used for identifying the syndrome of chronic gastritis, but it
also improves the syndrome diagnostic accuracy of chronic
gastritis.

4. Conclusions
To fully understand the characteristics of multilabel data of
TCM in syndrome diagnosis, feature selection was combined
with a multilabel learning algorithm.

Applying the REAL method extracts the relevant symp-
toms (signs) for each syndrome and improves the accuracy
of syndrome diagnosis in CG.

The study showed that the six finest symptoms (signs)
subsets agree with the theory and clinical practice of TCM. In addition, the study will serve as references for establishing
diagnostic criteria and a diagnostic model for CG and a
better guide for clinical practice. Further studies will focus
on building an intelligent diagnostic system for CG with
application of the method on biomedical data sets.

Acknowledgments
This work was supported by the National Natural Science
Foundation of China (Grant no. 30901897 and 81173199),
the Shanghai 3th Leading Academic Discipline Project
(Grant no. S30302), and the National Natural Science
Foundation of China (Grant no. 30701072.)

References
methods of traditional Chinese medicine,” Chinese Journal of
into regularity of syndrome classification for chronic atrophic
gastritis based on structural equation model,” Journal of
Nanjing University of Traditional Chinese Medicine, vol. 22, pp.
[3] Y. J. Bai, Q. G. Meng, H. B. Shen et al., “Study on non-
linear modelling of TCM syndrome of DN based on improved
BP neural network model,” Journal of Beijing University of


