INFORMATION STUDY

Symptom clustering in chronic gastritis based on spectral clustering

Wenhua Zhu, Zhaoxiang Fan, Guoping Liu, Jianjun Yan, Tao Zhong, Wu Zheng, Ruiqing Wang, Chunying Wang

Wenhua Zhu, Zhaoxiang Fan, Guoping Liu, Wu Zheng, Ruiqing Wang, Chunying Wang, Basic Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China
Jianjun Yan, Tao Zhong, Center for Mechatronic Engineering, East China University of Science and Technology, Shanghai 200237, China
Correspondence to: Associate Prof. Guoping Liu, Basic Medical College, Shanghai University of Traditional Chinese Medicine, Shanghai 201203, China. 13564133728@163.com; **Associate Prof. Jianjun Yan,** Center for Mechatronic Engineering, East China University of Science and Technology, Shanghai 200237, China. jjyan@ecust.edu.cn
Telephone: +86-13564133728; +86-13636507189

Abstract

OBJECTIVE: Apply spectral clustering to analyze the patterns of symptoms in patients with chronic gastritis (CG).

METHODS: Based on 919 CG subjects, we applied mutual information feature selection to choose the positively correlated symptoms with each pattern. Then, we used the Shi and Malik spectral clustering algorithm to select the top 20 correlated symptoms.

RESULTS: We ascertained the results of six patterns. There were three categories for the pattern of accumulation of damp heat in the spleen-stomach (0.00332). There were six categories for the pattern of dampness obstructing the spleen-stomach (0.02466). There were two categories for the pattern of spleen-stomach Qi deficiency (0.013 89). There were three categories for the pattern of spleen-stomach deficiency cold (0.009 15). There were five categories for the pattern of liver-Qi stagnation (0.01910). There were four categories for the pattern of stagnant heat in the liver-stomach (0.00585).

CONCLUSION: Most of the spectral clustering results of the symptoms of CG patterns were in accordance with clinical experience and Traditional Chinese Medicine theory. Most categories suggested the nature and/or location of the disease.

Key words: Gastritis; Cluster analysis; Pattern; Symptom complex

INTRODUCTION

Chronic gastritis (CG) involves chronic inflammation of the mucosa of the stomach. About 80%-90% of patients who undergo gastroscopy develop CG. Furthermore, the incidence of atrophic lesions increases gradually with age,[1-3] and CG is characterized by its long course and insidious development.[4]
Recently, it is found that Traditional Chinese Medi-

cine (TCM) is more efficacious than Western Medicine for the treatment of CG.[5] However, TCM is based on differentiation of disease patterns. In recent years, differentiation of disease patterns has been accepted as a method and applied widely to study various diseases.[6]

In this way, at first, symptoms that are closely related to each other are clustered as a category. Subsequently, the pathological meaning suggested by each category is analyzed in accordance with TCM theories to obtain previously unknown information. The category and its pathological meaning are named as "pattern factor". Finally, all the pattern factors are summed in one phrase: "pattern". During this process, the clustering of symptoms is vitally important .

Recently, a cluster algorithm was introduced into studies on pattern differentiation. For example, Li *et al*[7] applied frequency analysis, logistic regression and cluster analysis to cases of pneumonia and evaluated the various pattern factors involved, i.e., pathogen, nature and location of disease. Tang *et al*[8] applied R-type system cluster analysis to classify symptoms in cases of stomach ache, and discovered six patterns. They then employed factor analysis to calculate the contribution of symptoms. Chen *et al*[9] applied hierarchical clusters to classify symptoms in cases of gastritis and discovered three patterns. Studies have suggested that cluster analysis can be used to classify symptoms according to their degree of connection, and to choose representative ones (combined with TCM theories) to simplify the contents of the categories, thereby making the diagnosis simpler and convenient.[10]

The theories of spectral clustering originated from spectral graph partitioning. Compared with traditional cluster analysis, it is much more adept at clustering in the sample space of the discretional shape and converging to the optimal solution of the whole scenario.[11] In addition, it is not as sensitive to anomalous errors. Furthermore, the complexity of spectral clustering is lower than traditional cluster analysis if it is applied to higher-dimensional data such as textual data or image data. There is little information about the application of spectral clustering in syndrome clustering.

Based on the considerations stated above, we applied spectral clustering to study the relationship between the symptoms and pattern clustering of CG to provide insights into the pattern research of CG.

MATERIALS AND METHODS

The study protocol was approved by the Society of Medical Ethics connected to hospitals of Shanghai University of TCM. All patients provided written informed consent to participate in this study.

Subjects
Subjects with CG were recruited from the clinics, inpa-

tient departments, and gastroscopy rooms of the Digestive System Department of Longhua Hospital, Shuguang Hospital of the Shanghai University of Traditional Chinese Medicine, Xinhua Hospital, Putuo District Central Hospital, and the Shanghai Hospital of Traditional Chinese Medicine. All of these institutions are based in Shanghai, China. A total of 919 subjects [354 males (45 ± 15) years and 565 females (49 ± 13) years] were enrolled.

Inclusion and exclusion criteria
Patients who met the diagnostic criteria for CG and for analyses of TCM patterns were enrolled. Subjects who were mentally ill, had severe systemic diseases or who had difficulty describing their conditions were excluded.

Diagnostic criteria
The diagnosis was based on gastroscopy results, pathology results, and clinical performance according to the National Seminar on Consensus of CG formulated by the Digestive Diseases Branch of the Chinese Medical Association in 2006.[12]

Based on the Diagnosis Guidelines for Clinical Research of New Traditional Chinese Medicine[13] issued by the Ministry of Health, and Pattern Part of TCM Clinical Diagnosis and Treatment Terminology[14] issued by the China State Bureau of Technical Supervision, eight patterns of the TCM were assessed. These patterns were: (a) accumulation of damp heat in the spleen-stomach; (b) dampness obstructing the spleen-stomach; (c) *Qi* deficiency in the spleen-stomach; (d) cold deficiency in the spleen-stomach; (e) liver stagnation; (f) stagnant heat in the liver-stomach; (g) *Yin* deficiency in the stomach; (h) blood stasis in the stomach collateral.

Method for establishing the diagnostic scales of TCM studies
The research team undertaking this study comprised senior clinical experts on the digestive system, physicians, and clinical researchers based in Shanghai. The final diagnostic scales for TCM studies were drafted based on: previous experience in the production of such scales;[15] a wide range of literature on diseases of the spleen and stomach; Chinese scientific and technical articles selected by the Institute of Scientific and Technical Information of China over 15 years; reports on the frequency of symptoms associated with the TCM patterns of CG. The scales were also amended by two rounds of expert consultation and statistical tests. The scales comprised eight dimensions: cold or hot sweat; head, chest and abdomen; urine and stools; diet and taste; sleep; mood; female aspects; and contents of disease history, inspection, and palpation. Ultimately, >113 variables were included in the scales.

Investigation methods
Clear definitions of symptoms, specific methods, and