Saudi Journal of Biological Sciences 24 (2017) 1803–1809

Contents lists available at ScienceDirect

Saudi Journal of Biological Sciences

journal homepage: www.sciencedirect.com

Original article

Modeling analysis of the relationship between atherosclerosis and related inflammatory factors

Huidong Li [a], Pei Zhang [b], Shuaifang Yuan [b], Huiyuan Tian [b], Dandan Tian [b], Min Liu [b,*]

[a] Department of Hypertension, The Second Affiliated Hospital of Zhengzhou University, Henan Province, China
[b] Department of Hypertension, Henan Provincial People's Hospital, Henan Province, China

ARTICLE INFO

Article history:
Received 30 October 2017
Revised 7 November 2017
Accepted 7 November 2017
Available online 17 November 2017

Keywords:
Atherosclerosis
Inflammatory factor
Early diagnosis model

ABSTRACT

Objective: To establish early diagnosis model of inflammatory factors for atherosclerosis (AS), providing theoretical evidence for early detection of AS and development of plaques. *Methods:* Serum samples were collected to detect the inflammatory factors including CysC, Hcy, hs-CRP, UA, FIB, D-D, LP (a), IL-6, SAA, sCD40L and MDA. Using Logistic regression analysis, the inflammatory factors used for modeling were screened out, and then the AS early diagnosis models were established based on receiver operating characteristic (ROC) curve, support vector machine and BP neural network respectively. *Results:* No significant difference exists between the general materials of two groups. All 11 inflammatory factors had higher level in AS group than in control group. As shown in ROC curve, all inflammatory factors were helpful in AS diagnosis. In terms of sensitivity, UA ranked first (98) and FIB ranked last (55.5); in terms of specificity, UA ranked first (99) and FIB ranked last (78); in terms of area under the curve, UA and SAA ranked first (both were 0.995) and FIB ranked last (0.721). Based on Logistic regression equation, six factors were screened out, including Hcy, Hs-CRP, IL-6, D-D, CysC and MDA. According to classification, the final sixth steps had a prediction accuracy of 99%. When six inflammatory factors included in Logistic regression equation were detected jointly, the sensitivity, specificity and area under the curve were 57%, 97% and 0.821 respectively, while those of the model excluding D-D were 64%, 90% and 0.828, generally superior to results of joint detection including six factors. The ROC curve based on Hcy, Hs-CRP and MDA had a sensitivity of 87%, a specificity of 94% and an area under the curve of 0.869, being inferior to those of the ROC curve based on IL-6, D-D and Cys C, which were 87%, 92% and 0.936 respectively. The accuracy of SVM-AS diagnosis model and BP neural network model were 82.5% and 77.5% respectively. *Conclusion:* All 11 inflammatory factors are valuable in AS diagnosis. AS early diagnosis models based on Logistic regression analysis, ROC curve, support vector machine and BP neural network possess diagnostic value and can provide reference for clinical diagnosis.

1. Introduction

At present, cardiovascular disease has become the primary cause of death in the world, of which, atherosclerosis is the main cause of cardiac death. Deaths in atherosclerosis in Europe and the United States account for 1/3 of the total deaths. Domestically, atherosclerosis morbidity and mortality rate grows rapidly, greatly threatening life. Atherosclerosis (AS) is a type of arteriosclerosis, also a most important type of vascular disease. AS is often accompanied by high blood pressure, hypercholesterolemia or diabetes, etc. (Blood Lipids and Atherosclerosis Group, 2017; Rezaei-Hachesu et al., 2017), more prevalent in cerebral arteries, coronary artery, aorta. There are many factors influencing the occurrence and development of atherosclerotic plaques, including lipid infiltration, damage to mononuclear cells, arterial endothelial cells, foam cells as a result of macrophage phagocytosis of lipid, repair reaction after vascular injury, etc.

Recent years see increasing popularity of bioinformatics in data mining. Data mining is the process of digging out effective, potentially useful, novel and eventually understandable patterns from excessive data, which can also be understood as extracting or

* Corresponding author at: Department of Hypertension, Henan Provincial People's Hospital, No. 7, Weiwu Road, Zhengzhou, Henan 450000, China.
E-mail address: liumin136@126.com (M. Liu).
Peer review under responsibility of King Saud University.

Production and hosting by Elsevier

https://doi.org/10.1016/j.sjbs.2017.11.016
1319-562X/

digging knowledge from excessive data. In general, clinical medical data feature diversity, redundancy repeatability, complexity, temporal priority and non-normality. Via data mining, we can extract valuable information from complex medical data to help with clinical decision making. The auxiliary diagnostic models of data mining including Logistic Regression (LR), Support Vector Machine (SVM), Artificial Neural Network (ANN), Decision Tree (DT), Bayes are more and more used in medical diagnosis (Tseng et al., 2017; Wildenberg et al., 2017). Support vector machine is a category of computer-aided diagnosis. As an auxiliary diagnostic tool, it cannot yet completely replace the clinician's diagnosis (Cinelli et al., 2017; Caixinha et al., 2016), but its auxiliary diagnostic value has been recognized.

In this study based on the development of atherosclerotic plaques, such as changes in coagulation and immune levels, oxidative stress and inflammation, plasma cystatin C (Cys), homocysteine (Hcy), D-dimer (D-D), hypersensitive C reactive protein (hs-CRP), malondialdehyde (MDA), uric acid (UA), interleukin-6 (IL-6), soluble CD40 ligand (sCD40L), lipoprotein (a) [LP (a)], fibrinogen (FIB) and serum amyloid A (SAA) were selected from the inflammatory factors that could reflect these changes, to detect the correlation with AS and evaluate the application value in AS diagnosis. The application value of the biochemical markers in AS clinical diagnosis was evaluated based on support vector machine, BP neural network, Logistic regression analysis and receiver operating characteristic (ROC), to establish diagnostic model of serum markers in AS early diagnosis, assist doctors' diagnosis, improve the diagnosis rate, thereby laying the theoretical basis for early clinical detection of atherosclerosis.

2. Materials and methods

2.1. General data

The 200 patients with atherosclerosis who were hospitalized in our hospital from 2012 to 2015 were selected as the experimental group. With 42 males and 18 females, the group was aged 43–75 years (mean age 63.5 years). Another 100 healthy people who had physical examination in our hospital for the same period were selected as the control group. With 43 males and 17 females, the group was aged 43–76 years (mean age 64.7 years). Exclusion criteria: immune disease, with acute and chronic infection evidence, tumor, recent surgery or trauma, chronic connective tissue disease and valvular disease, atrial fibrillation, severe renal insufficiency (serum creatinine >120 Lmol/L), hyperthyroidism, iodine allergy. All the subjects' age, sex, history of hypertension, diabetes, alcohol and tobacco were recorded. This study has received written consent from all patients and follows the Helsinki Declaration and other bioethical principles.

2.2. Serum collection

All subjects had 12 h fasting since the previous night. Afterwards, 10 ml venous blood was collected in the morning and placed in EDTA anticoagulant tube. The sample underwent centrifugation within 2 h at speed of 3000 r/min for 10 min. The serum sample was stored at −20 °C.

2.3. Indicator detection

The contents of LDL-C, HDL-C, TC, TG, UA, Hcy, CysC and hs-CRP were determined by automatic biochemical analyzer (Hitachi-7100). The related reagents were provided by Shanghai Diasys Company. The volumes of LP (a) and D-dimer were determined by immunoturbidimetry. LP (a) was determined by immune scatter turbidity and IMMAGE dual-ray rate turbidity analysis system. The system and related reagents were from Beektnan Coulter. D-dimer was determined by latex immunoturbidimetry using fully automated coagulation analyzer (Sysmex CA-550). IL-6, SAA, sCD40L were detected by enzyme-linked immunosorbent assay (ELISA) (with the kit provided by Shanghai Hengyuan Biotechnology Co., Ltd.). MDA: follow instructions on MDA kit (provided by Shanghai Jining Biotechnology) using colorimetry.

2.4. Model establishment method

Diagnostic model establishment based on Logistic regression analysis: the above 11 inflammatory factor detection levels of AS experimental group and the healthy control group was subject to binary variable assignment, with 0 and 1 for normal and abnormal inflammatory factor detection level respectively. With the 12 indicators as concomitant variables, with pathological diagnosis result of atherosclerosis (AS patient = 1, health group = 0) as dependent variable, make gradual logistic regression analysis with forward method to screen out the biochemical markers used to determine presence of atherosclerosis, and ultimately obtain the modeling indicators. With detection level of the modeling indicators as test variable, and pathological diagnosis result as state variable, formulate the ROC curve. After data entry in SPSS 17.0, with detected level of the 11 inflammatory factors as test variable, and pathological diagnosis result as state variable, formulate separate ROC curve, and evaluate the diagnostic value for atherosclerosis based on area under the curve (AUC).

Diagnostic model establishment based on support vector machine: the collected 300 cases of data were normalized, with patient marked as 1 and healthy one marked as 0. Randomly select 180 out of 200 cases of atherosclerosis and 80 out of 100 cases of healthy people as training set sample to be input to support vector machine for training. The remaining 20 cases of atherosclerosis and 20 healthy subjects were input to support vector machine network after training as test set sample. The discrimination accuracy can be obtained after comparing the discrimination results (1 or 0) with the object.

Diagnostic model establishment based on BP neural network: the 6 parameters of Hcy, IL-6, Hs-CRP, DD, CysC and MDA were incorporated in the study to establish BP neural network model. The case data were randomly divided into training set and test set. The training set and test set data were normalized before input in the network.

2.5. Statistical analysis

All data in this study were analyzed by SPSS l7.0 software. $P < .05$ indicates statistically significant difference.

3. Results

3.1. General clinical data

The general data of the two groups of subjects include age, body mass index, sex, hypertension, diabetes, alcohol consumption, smoking, triglyceride, total cholesterol, low density lipoprotein cholesterol and high density lipoprotein cholesterol. These indicators have no significant differences (as shown in Table 1), $P > .05$.

3.2. Content detection result of each inflammatory factor

The inflammatory factor test results of atherosclerosis patients and healthy control group (the two totaling 300 cases) are shown in Table 2. The 11 inflammatory factors are Cys C, Hcy, DD,