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

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

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,

**Table 1**

General data and lipid level comparison between atherosclerosis group and the control group.

| Item | Control group | Experimental group | F($\chi^2$) | P |
|------|---------------|--------------------|------------|----|
| Case number | 100 | 200 | | |
| Age | 60.77 ± 8.05 | 62.67 ± 8.63 | 1.04 | .07 |
| Body mass index | 23.31 ± 2.08 | 23 ± 2.23 | 0.30 | .26 |
| Sex (male/female) | 58/42 | 122/78 | 0.25 | .35 |
| Hypertension | 58(58.0%) | 142(71.0%) | 1.40 | .24 |
| Type 2 diabetes | 32(32.0%) | 96(48.0%) | 2.81 | .09 |
| Alcohol consumption | 33(33.0%) | 92(46.0%) | 2.77 | .10 |
| Smoking history | 29(29.0%) | 90(45.0%) | 2.37 | .12 |
| TG (mmol/L) | 1.37 ± 0.66 | 1.53 ± 0.72 | 0.83 | .06 |
| TC (mmol/L) | 4.65 ± 1.13 | 4.90 ± 1.12 | 1.07 | .07 |
| LDL-C (mmol/L) | 2.35 ± 0.71 | 2.50 ± 0.69 | 0.00 | .08 |
| HDL-C (mmol/L) | 1.27 ± 0.41 | 1.20 ± 0.35 | 3.49 | .13 |

**Table 2**

Test results of 11 inflammatory factors in atherosclerosis ($\bar{x} \pm s$).

| Detection index | Control group (n = 100) | Experimental group (n = 200) | F | P |
|-----------------|-------------------------|------------------------------|-----|----|
| Hcy (umol/L) | 9.08 ± 3.16 | 19.66 ± 8.36 | 55.15 | .00 |
| IL-6 (pg/mL) | 112.57 ± 21.86 | 152.09 ± 28.75 | 7.08 | .00 |
| Hs-CRP (mg/L) | 1.47 ± 0.77 | 4.23 ± 1.98 | 59.81 | .00 |
| D-D (mg/L) | 0.88 ± 0.39 | 1.97 ± 1.52 | 83.76 | .00 |
| CysC (mg/L) | 0.86 ± 0.19 | 1.29 ± 0.35 | 29.60 | .00 |
| UA (umol/L) | 221.42 ± 23.74 | 319.63 ± 24.46 | 0.09 | .00 |
| SAA (mg/L) | 0.19 ± 0.24 | 4.69 ± 2.48 | 157.22 | .00 |
| LP(a) (mg/L) | 127.4 ± 51.04 | 319.35 ± 129.95 | 69.26 | .00 |
| MDA (ng/ml) | 3.45 ± 0.82 | 5.46 ± 0.93 | 0.88 | .00 |
| FIB (g/L) | 2.93 ± 0.76 | 3.59 ± 0.84 | 1.46 | .00 |
| sCD40L (ng/mL) | 4.49 ± 0.73 | 5.89 ± 1.05 | 10.39 | .00 |

Note: P < .05 indicates a significant difference.

hs-CRP, UA, MDA, IL-6, FIB, sCD40L, LP (a), SAA. As can be seen from Table 2, level of the 11 inflammatory factors in the atherosclerosis group is significantly higher than that in the healthy control group ($P$ < .05).

### 3.3. ROC curve

ROC curve which is often used to evaluate the pros and cons of a binary classifier is generally above the straight line of y = x. The closer to the upper left corner the ROC curve is, the closer to 1 the area is, and the better the classification effect is. Diagnostic value of these inflammatory factors for atherosclerosis is evaluated with ROC curve. Figs. 1, 2 and Table 3 show certain diagnostic value of each inflammatory factor for atherosclerosis. In terms of sensitivity, UA ranks the first with 98, FIB ranks the last with 55.5; in terms of specificity, UA ranks the first with 99, FIB ranks the last with 78; in terms of area under the curve, UA and SAA are the highest with 0.995, FIB is the lowest with 0.721.
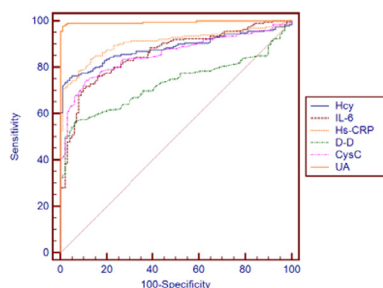


**Fig. 1.** ROC curve of various inflammatory factors (Hcy, IL-6, Hs-CRP, D-D, CysC, UA) versus atherosclerosis.
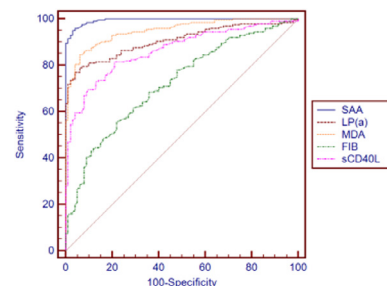


**Fig. 2.** ROC curve of various inflammatory factors (SAA, LP(a), MDA, FIB, sCD40L) versus atherosclerosis.

**Table 3**

Diagnostic value of each inflammatory factor for atherosclerosis.

| | Sensitivity | Specificity | AUC |
|------|-------------|-------------|------|
| Hcy | 76.5 | 95 | 0.883 |
| IL-6 | 72 | 89 | 0.859 |
| Hs-CRP | 78.5 | 92 | 0.908 |
| D-D | 56.5 | 94 | 0.732 |
| CysC | 74.5 | 89 | 0.857 |
| UA | 98 | 99 | 0.995 |
| SAA | 96 | 96 | 0.995 |
| LP(a) | 79 | 94 | 0.906 |
| MDA | 84.5 | 94 | 0.95 |
| FIB | 55.5 | 78 | 0.721 |
| sCD40L | 81.5 | 79 | 0.865 |

### 3.4. Logistic regression analysis results

With the 11 parameters of CIS C, Hcy, DD, hs-CRP, UA, MDA, IL-6, FIB, sCD40L, LP (a) and SAA as concomitant variables, with AS pathological diagnosis result as dependent variable, gradual Logistic regression analysis was made with forward method, with results shown in Tables 4 and 5. As can be seen from Tables 4 and 5, six independent variables of the 12 indicators are deleted, and six independent variables are selected for the Logistic regression equation, namely, Hcy, Hs-CRP, IL-6, D-D, CysC and MDA. The partial regression coefficients are 0.275, 1.202, 0.065, 1.989, 9.724 and 3.407, respectively. The corresponding P values are 0.032, 0.035, 0.003, 0.029, 0.003 and 0.012, respectively, which are less than 0.05 with statistical significance. The classification table (Table 5) shows classification prediction of atherosclerosis in each step. The first step has a diagnostic prediction accuracy of 88.7%, while accuracy of the second, third, fourth, fifth and sixth steps is 93%, 96%, 97.7%, 98.3% and 99% respectively.

### 3.5. ROC Curve Analysis of Logistic Regression Model for Individual and Joint Detection of 6 Indicators

ROC curve which is often used to evaluate the pros and cons of a binary classifier is generally above the straight line of y = x. The closer to the upper left corner the ROC curve is, the closer to 1 the area is, and the better the classification effect is. The six inflammatory factors incorporated in Logistic regression analysis equation are subject to individual and joint ROC curve analysis, with results shown in Figs. 3, 4 and Table 6. In individual detection, AUC of MDA is the highest, specificity of Hcy is the strongest, and sensitivity of D-D is the worst. In joint detection of the 6 inflammatory factors, sensitivity is 57%, specificity is 97% and the area under the curve is 0.821. The model sensitivity is low which may be because of the affect of the too poor D-D sensitivity. Thus, ROC curve analysis was made after D-D removal, and then sensitivity increased by 64%, specificity reduced to 90%, the area under the curve was

**Table 4**
Variables in the Logistic Regression Equation.

|        |          | B       | S.E.   | Wals   | df | Sig. | Exp (B)    |
|--------|----------|---------|--------|--------|----|------|------------|
| Step 1[a] | MDA      | 2.659   | .319   | 69.372 | 1  | .000 | 14.284     |
|        | constant | -11.099 | 1.388  | 63.964 | 1  | .000 | .000       |
| Step 2[b] | Cys C    | 6.295   | 1.142  | 30.398 | 1  | .000 | 541.826    |
|        | MDA      | 3.105   | .471   | 43.484 | 1  | .000 | 22.319     |
|        | constant | -19.687 | 2.853  | 47.607 | 1  | .000 | .000       |
| Step 3[c] | Cys C    | 6.751   | 1.551  | 18.956 | 1  | .000 | 855.157    |
|        | IL-6     | .072    | .016   | 20.909 | 1  | .000 | 1.074      |
|        | MDA      | 3.311   | .600   | 30.406 | 1  | .000 | 27.403     |
|        | constant | -30.028 | 4.885  | 37.779 | 1  | .000 | .000       |
| Step 4[d] | Hcy      | .318    | .085   | 13.990 | 1  | .000 | 1.374      |
|        | Cys C    | 7.186   | 1.918  | 14.038 | 1  | .000 | 1320.170   |
|        | IL-6     | .077    | .019   | 16.048 | 1  | .000 | 1.080      |
|        | MDA      | 3.554   | .794   | 20.015 | 1  | .000 | 34.953     |
|        | constant | -36.178 | 7.012  | 26.618 | 1  | .000 | .000       |
| Step 5[e] | Hcy      | .270    | .098   | 7.623  | 1  | .006 | 1.311      |
|        | D-D      | 2.238   | .865   | 6.692  | 1  | .010 | 9.375      |
|        | Cys C    | 8.826   | 2.819  | 9.802  | 1  | .002 | 6809.635   |
|        | IL-6     | .068    | .019   | 12.456 | 1  | .000 | 1.070      |
|        | MDA      | 4.198   | 1.323  | 10.075 | 1  | .002 | 66.568     |
|        | constant | -41.839 | 10.282 | 16.559 | 1  | .000 | .000       |
| Step 6[f] | Hcy      | .275    | .128   | 4.584  | 1  | .032 | 1.316      |
|        | Hs-CRP   | 1.202   | .570   | 4.441  | 1  | .035 | 3.325      |
|        | D-D      | 1.989   | .913   | 4.743  | 1  | .029 | 7.309      |
|        | Cys C    | 9.724   | 3.312  | 8.622  | 1  | .003 | 16718.102  |
|        | IL-6     | .065    | .022   | 8.592  | 1  | .003 | 1.067      |
|        | MDA      | 3.407   | 1.351  | 6.363  | 1  | .012 | 30.167     |
|        | constant | -41.310 | 11.015 | 14.066 | 1  | .000 | .000       |

[a] The variable entered in step 1: MDA.
[b] The variable entered in step 2: Cys C.
[c] The variable entered in step 3: IL-6.
[d] The variable entered in step 4: Hcy.
[e] The variable entered in step 5: D-D.
[f] The variable entered in step 6: Hs-CRP.

**Table 5**
Classification table[a].

|        | Observed         |                    | Predicted     |                    |                       |
|--------|------------------|--------------------|---------------|--------------------|-----------------------|
|        |                  |                    | Atherosclerosis |                  | Percentage correction |
|        |                  |                    | Control group | Experimental group |                       |
| Step 1 | Atherosclerosis  | Control group      | 80            | 20                 | 80.0                  |
|        |                  | Experimental group | 14            | 186                | 93.0                  |
|        | Total percentage |                    |               |                    | 88.7                  |
| Step 2 | Atherosclerosis  | Control group      | 87            | 13                 | 87.0                  |
|        |                  | Experimental group | 8             | 192                | 96.0                  |
|        | Total percentage |                    |               |                    | 93.0                  |
| Step 3 | Atherosclerosis  | Control group      | 94            | 6                  | 94.0                  |
|        |                  | Experimental group | 6             | 194                | 97.0                  |
|        | Total percentage |                    |               |                    | 96.0                  |
| Step 4 | Atherosclerosis  | Control group      | 97            | 3                  | 97.0                  |
|        |                  | Experimental group | 4             | 196                | 98.0                  |
|        | Total percentage |                    |               |                    | 97.7                  |
| Step 5 | Atherosclerosis  | Control group      | 97            | 3                  | 97.0                  |
|        |                  | Experimental group | 2             | 198                | 99.0                  |
|        | Total percentage |                    |               |                    | 98.3                  |
| Step 6 | Atherosclerosis  | Control group      | 99            | 1                  | 99.0                  |
|        |                  | Experimental group | 2             | 198                | 99.0                  |
|        | Total percentage |                    |               |                    | 99.0                  |

[a] Cut value is .500.

0.828, which was superior to that of joint detection in overall. Considering the unsatisfactory result, the three biochemical markers Hcy, Hs_CRP, MDA with superior effect and the three inflammatory factors, IL_6, D_D and Cys C with poor effect were respectively joined for ROC curve analysis. According to the result, in the former joint detection, sensitivity, specificity and area under the curve were 67%, 94% and 0.869, respectively, which was inferior to 87% sensitivity, 92% specificity and 0.936 area under the curve in the latter joint detection. The specific reason may be the interaction between inflammatory factors, which needs further study.
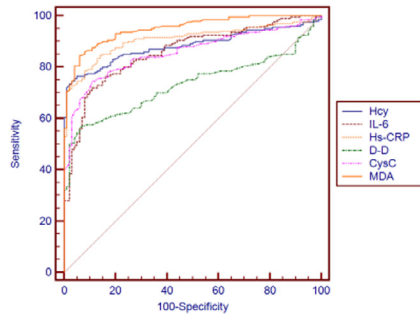
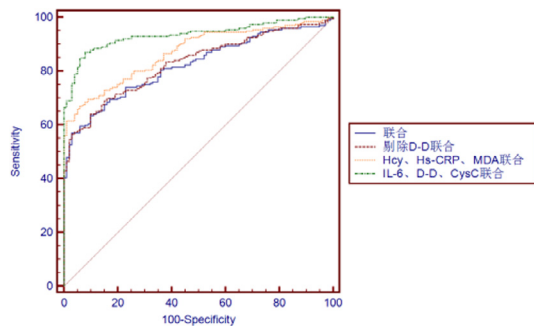**Fig. 3.** ROC curve of individual detection versus atherosclerosis.



**Fig. 4.** ROC curve of several joint detections versus atherosclerosis.

### 3.6. Diagnostic model establishment based on SVM

SVM atherosclerosis diagnostic model was established by incorporating six inflammatory factors of Hcy, IL-6, Hs-CRP, D-D, CysC, MDA, as shown in Fig. 3. The empty circle represents the target output, "∗" is the actual simulation output of SVM. As can be seen from the figure, accuracy of the diagnostic model is 82.5% (Fig. 5).

### 3.7. Diagnostic Model Establishment based on BP Neural Network

As can be seen from Fig. 6 parameters of Hcy, IL-6, Hs-CRP, D-D, CysC and MDA are first incorporated in establishment of the neural network. In this study, 260 out of 300 samples were selected as training samples for the training set. The remaining 40 samples constitute the test set. In the training process, a few parameters should be set: incentive function, training function, number of hidden layers, number of hidden layer nodes, number of network output nodes, training frequency threshold and accuracy. Different parameter combinations will achieve different effects. In this experiment, several parameters such as incentive function, conversion function and test function are crossed and combined, and the best combination is selected by training of training samples with various combinations. The results show that the optimal combination is tansig for incentive function, purelin for conversion function, trainlm for training function, 1 for number of hidden layers, 10 for number of hidden layer nodes, 1 for number of output nodes, 10,000 for training frequency threshold and 0.001 for accuracy. The accuracy rate of the model was 77.5% and the misdiagnosis rate was 22.5%.

## 4. Discussions

With the development of natural science, human beings have increasingly deeper understanding towards AS occurrence and development. At present, most scholars believe that AS is very

**Table 6**
Diagnostic value of individual and joint detection for atherosclerosis.

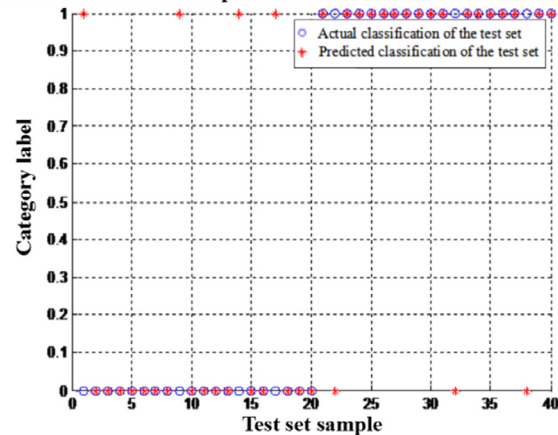|  | Sensitivity | Specificity | AUC |
|---|---|---|---|
| Hcy | 76.5 | 95 | 0.883 |
| IL-6 | 72 | 89 | 0.859 |
| Hs-CRP | 78.5 | 92 | 0.908 |
| D-D | 56.5 | 94 | 0.732 |
| CysC | 74.5 | 89 | 0.857 |
| MDA | 84.5 | 94 | 0.95 |
| Joint detection | 57 | 97 | 0.821 |
| Joint detection after D-D removal | 64 | 90 | 0.828 |
| Hcy, Hs-CRP, MDA joint detection | 67 | 94 | 0.869 |
| IL-6, D-D, CysC joint detection | 87 | 92 | 0.936 |



**Fig. 5.** Accuracy test results of SVM model established by incorporating six inflammatory factors.

likely to be an inflammatory disease. The theoretical basis is that plaque instability can be caused by the activation of inflammatory response, thereby AS patients have acute cerebral infarction. Studies have confirmed that chronic inflammation plays a vital role in the process of AS occurrence and development (Ridler et al., 2000).

This study shows that inflammatory factors CysC, Hcy, hs-CRP, UA, FIB, DD, LP (a), IL-6, SAA, sCD40L, MDA have significant correlation with atherosclerosis, which can serve as sensitivity indicator of AS diagnosis. This is also consistent with many previous studies (Kwon et al., 2017; Chen, 2015; Sun and Guo, 2017). High sensitivity C-reactive protein (hs-CRP) is a highly sensitive inflammatory marker, which can accurately detect low concentration CRP in serum. Clinical studies have shown that hs-CRP is an important risk factor for AS (Sara et al., 2017), which can effectively predict inflammatory responses as independent risk factor of cardiovascular disease. CRP can also increase the expression of adhesion factors, which promotes vascular endothelial cell proliferation as an inseparable factor in AS occurrence and development (Xiao et al., 2016). Lipoprotein (a) (Lp (a)), as an independent macromolecular protein with specific antigenicity, can interfere with lipid metabolism and fibrinolytic system, thus playing a vital role in development of cardiovascular disease thrombosis and AS (Yang et al., 2017). Clinically, metabolic site and metabolic mechanisms of LP (a) in vivo remain unclear, and racial genetic factors have a greater impact on them. A number of studies have shown that atherosclerotic disease is closely related to LP (a), and LP (a) rise is an independent risk factor for cardiovascular events (Fan et al., 2017; Xu et al., 2017).

The sulfur-containing amino acids in the human body produce an important intermediate metabolite, i.e. homocysteine (Hcy). Studies have shown that elevated Hcy level in plasma is closely
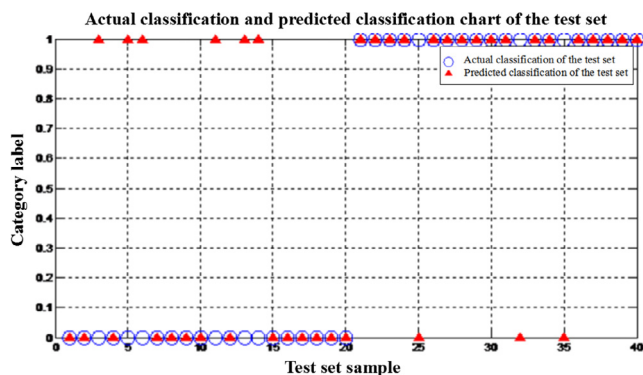
**Fig. 6.** Output results after BP neural network training.

related to atherosclerosis. The main mechanism is that elevated Hcy level can lead to endothelial cell damage, and thus promote vascular smooth muscle proliferation and platelet aggregation (Fu et al., 2015; Gurda et al., 2015). Uric acid (UA) is an inflammatory substance that promotes platelet aggregation and thrombosis (Sharaf El Din et al., 2017). Uric acid can lead to vascular diastolic dysfunction, increased inflammatory cells, lipid deposition in the arterial intima, and vascular intima damage, causing atherosclerosis exacerbation (Hasic et al., 2017). Serum amyloid A (SAA) is a group of polymorphic proteins that are expressed under indirect stimulation of I-1, IL-6 and TNF-α cytokines. Studies have shown that SAA is a very sensitive inflammatory marker (Yuan et al., 2016). IL-6 is a class of peptide-like cytokines produced by T lymphocytes and mononuclear macrophages, which have immunoregulatory functions. Some scholars have suggested that cytokine IL-6 is closely related to occurrence and progression of carotid atherosclerosis in patients with hypertension (Wan et al., 2017). Other studies have shown that serum IL-6 levels are positively correlated with the size of cerebral infarction volume (Zeng and Lu, 2016). D-dimer is the most simple and minimal product of thrombus degradation by fibrinolytic action on coagulated thrombus. D-dimer can be detected by immunological principles by simple and accurate method; so it can serve as the only exact and particularly sensitive specific marker reflecting hypercoagulable state and secondary fibrinolytic activity in vivo (Wang et al., 2017). As the research progresses, D-dimer is also increasingly used in patients with atherosclerosis (Takamura et al., 2017). Cystatin C (CysC) is an alkaline non-saccharified protein. Most investigators believe that CysC is closely associated with cardiovascular and cerebrovascular disease, and its unbalanced expression will lead to atherosclerosis and aneurysms (Li et al., 2016). Kral et al. (2016) gave clear evidence for CysC's relationship with AS. That is, the substance is closely related to stability and regression of atherosclerotic plaques which involves overexpression of cathepsin and low expression of corresponding inhibitors.

CD40-CD40L system is the hub of immune response and inflammatory response. Experimental results (Geng et al., 2017) show excessive CD40 and CD40L secretion in AS patients, which is more severe in patients with plaques prone to rupture. CD40L can promote tissue factor expression and thrombosis within the plaque. Studies have also shown that blocking CD40L in mice after vascular endothelial injury not only promotes plaque stability in progression, but also causes less formation of AS plaque (Hueso et al., 2016). Fibrinogen (FIB) is a key coagulation factor during coagulation reaction. According to the study (Yang et al., 2017), abundant FIB and its degradation products exist in atherosclerotic plaques, which plays an important role in stimulating smooth muscle cell proliferation and migration, while promoting low-density lipoprotein adsorption in the vascular intima, thereby increasing lipid

aggregation in plaques. Malondialdehyde (MDA) is mainly present in low density lipoprotein cholesterol (LDL-C). It can act on the lipid and produce lipid peroxidation products. Lipid peroxidation is one of the initiating links of endothelial dysfunction, and MDA levels can exactly reflect the severity of lipid peroxidation (LP) injury, which is also closely related to atherosclerosis (Jia et al., 2017).

Since the 1990s, tens of thousands of biomedical data have sprung up rapidly along with progress of various genome sequencing programs and "precision medical" programs. Bioinformatics data analysis and processing methods are also increasingly important in the processing of a large number of medical test data. Logistic regression has three main purposes: risk factor search, predictions and judgments. As one of the most widely used statistical methods in medicine, Logistic regression is the most commonly used models in prediction of complications, such as prediction of risk in surgical complications of gastric cancer (Zhou et al., 2016), prostate biopsy factors and prostate cancer prediction (Li et al., 2015). In the present study, by Logistic regression analysis, 6 biochemical indexes of Hcy, IL-6, Hs-CRP, DD, CysC and MDA were incorporated in the equation, with respective partial regression coefficients at 0.275, 1.202, 0.065, 1.989, 9.724 and 3.407. With corresponding P value less than 0.05, there is statistical significance. The receiver operating characteristic curve (ROC) analysis was based on a series of different binary methods. ROC curve of the prediction model was plotted by Medcalc software, with area under each curve (AUC) calculated. The largest area under the ROC curve indicates that the test has the best diagnostic value. Gu et al. (2017) used ROC curve to evaluate diagnostic value of free fatty acids (FFA) for coronary heart disease (CHF). Zong et al. (2015) used Logistic regression and ROC curve to analyze the diagnostic value of the three serum markers in primary liver cancer (PHC), screening out two markers with higher diagnostic value. In this study, the six biochemical markers incorporated into logistic equation were subject to individual and joint ROC curve analysis. In individual detection, MDA had the largest area under the curve, Hcy had the strongest specificity, and D-D had the worst sensitivity. In joint detection of the 6 biochemical markers, sensitivity was 57%, specificity was 97%, and the area under the curve was 0.821. In ROC curve analysis after D-D removal, sensitivity improved by 64%, specificity reduced to 90%, and area under the curve was 0.828, which was superior to those of joint detection in overall. In respective joint ROC curve analysis of Hcy, Hs_CRP, MDA and IL_6, D_D, Cys C, sensitivity, specificity and area under the curve of the former joint detection were 67%, 94% and 0.869, respectively, which was inferior to 87% sensitivity, 92% specificity and 0.936 area under the curve of the latter joint detection. Support vector machine (SVM), as the category of informatics, is the best theory of small sample learning. Widely used in the field of intelligent medical data analysis, it represents a hot topic in the current intelligent medical diagnosis research (Wang et al., 2016). Zhao et al. (2017) proposed a method of coronary artery lesion detection based on support vector machine, and improved the accuracy of algorithm recognition by using coronary surface resampling and feature selection method based on maximum mutual information. BP (Back Propagation) neural network is a multi-layer feedforward network. Trained in accordance with error reverse propagation algorithm, it can solve multi-layer neural network learning problem. In recent years, BP neural networks have been widely used in many medical fields, such as neonatal birth, disease diagnosis, disease prognosis, risk assessment (Kar and Majumder, 2017; Jiang et al., 2015). In this study, the six biochemical markers of Hcy, IL-6, Hs-CRP, DD, CysC and MDA were analyzed by SVM modeling and BP neural network. The accuracy of SVM diagnostic model was 82.5% and that of BP neural network was 77.5%.

In summary, all the selected 11 inflammatory factors have diagnostic value for AS, but the selected inflammatory factors are a result of human decision, so we failed to incorporate all factors influencing AS. Further in-depth mate analysis is required in this aspect to avoid omission of some important factors. In addition, in modeling analysis of the selected inflammatory factors by using various intelligent algorithms, the number of samples is limited, which will have some effect on the results and can only explain partial circumstances. Therefore, modeling analysis is needed by increasing the sample size.

## Fund

## References

Blood Lipids and Atherosclerosis Group. Chinese Society of Integrative Medicine Cardiovascular Disease Committee. Integrative medicine expert consensus on atherosclerosis. Chin. Gen. Pract. 20(5), 507–511.

Caixinha, M., Amaro, J., Santos, M., et al., 2016. In-vivo automatic nuclear cataract detection and classification in an animal model by ultrasounds. IEEE Trans. Biomed. Eng. 63 (11), 2326–2335.

Chen, X.Z., 2015. Prediction value of serum inflammatory factors in coronary heart disease plaque vulnerability. Chin. Clin. Doctor 43 (9), 41–43.

Cinelli, M., Sun, Y., Best, K., et al., 2017. Feature selection using a one dimensional naive Bayes' classifier increases the accuracy of support vector machine classification of CDR3 repertoires. Bioinformatics 33 (7), 951–955.

Fan, Y., Hu, J.S., Guo, F., et al., 2017. Lipoprotein(a) as a predictor of poor collateral circulation in patients with chronic stable coronary heart disease. Braz. J. Med. Biol. Res = Revista brasileira de pesquisas medicas e biologicas 50 (8), e5979.

Fu, Z., Qian, G., Xue, H., et al., 2015. Hyperhomocysteinemia is an independent predictor of long-term clinical outcomes in Chinese octogenarians with acute coronary syndrome. Clin. Interventions Aging 10, 1467–1474.

Geng, J.M., Wan, C.W., Le, C.Y., et al., 2017. Relationship between CD40L levels and lesion structural changes in human coronary atherosclerotic lesions. Chin. J. Atherosclerosis 25 (4), 355–359.

Gu, K.P., Wang, Y.P., Yu, W., et al., 2017. Application of ROC curve to analyze the clinical value of free fatty acids for patients with coronary heart disease. Lab. Med. 32 (5), 367–369.

Gurda, D., Handschuh, L., Kotkowiak, W., et al., 2015. Homocysteine thiolactone and N-homocysteinylated protein induce pro-atherogenic changes in gene expression in human vascular endothelial cells. Amino acids 47 (7), 1319–1339.

Hasic, S., Kadic, D., Kiseljakovic, E., 2017. Serum uric acid could differentiate acute myocardial infarction and unstable angina pectoris in hyperuricemic acute coronary syndrome patients. Med. Arch. (Sarajevo, Bosnia and Herzegovina) 71 (2), 115–118.

Hueso, M., De Ramon, L., Navarro, E., et al., 2016. Silencing of CD40 in vivo reduces progression of experimental atherogenesis through an NF-kappaB/miR-125b axis and reveals new potential mediators in the pathogenesis of atherosclerosis. Atherosclerosis 255, 80–89.

Jia, J.F., Liu, Y.M., Yang, W.D., et al., 2017. Relationship between serum HCY levels and oxidative stress in patients with essential hypertension complicated with carotid atherosclerosis. Shandong Med. J. 57 (10), 80–81.

Jiang, Y.P., Zhang, Y., Jiang, N., et al., 2015. Colorectal cancer early warning by tumor markers combined with artificial neural network. Chin. J. Health Lab. Technol. 3, 371–373.

Kar, S., Majumder, D.D., 2017. A mathematical theory of shape and neuro-fuzzy methodology-based diagnostic analysis: a comparative study on early detection and treatment planning of brain cancer. Int. J. Clin. Oncol. 22 (4), 667–681.

Kral, A., Kovarnik, T., Vanickova, Z., et al., 2016. Cystatin C is associated with the extent and characteristics of coronary atherosclerosis in patients with preserved renal function. Folia Biologica 62 (6), 225–234.

Kwon, O., Kang, S.J., Kang, S.H., et al., 2017. Relationship between serum inflammatory marker levels and the dynamic changes in coronary plaque characteristics after statin therapy. Circ. Cardiovasc. Imaging 10 (7), e005934.

Li, W., Sultana, N., Siraj, N., et al., 2016. Autophagy dysfunction and regulatory cystatin C in macrophage death of atherosclerosis. J. Cell Mol. Med. 20 (9), 1664–1672.

Li, Y., Tang, Z., Qi, L., 2015. Analysis of influential factors for prostate biopsy and establishment of logistic regression model for prostate cancer. Zhong nan da xue xue bao Yi xue ban = J. Cent. S. Univ. Med. Sci. 40 (6), 651–656.

Rezaei-Hachesu, P., Oliyaee, A., Safaie, N., et al., 2017. Comparison of coronary artery disease guidelines with extracted knowledge from data mining. J. Cardiovasc. Thoracic Res. 9 (2), 95–101.

Ridler, P.M., Henekerrs, C.H., Buring, J.E., et al., 2000. C- reactive protein and other markers of inflammation in the prediction of cardiovascular disease in women. N. Engl. J. Med. 342 (12), 836–843.

Sara, J.D.S., Prasad, M., Zhang, M., et al., 2017. High-sensitivity C-reactive protein is an independent marker of abnormal coronary vasoreactivity in patients with non-obstructive coronary artery disease. Am. Heart J. 190, 1–11.

Sharaf El Din, U.A.A., Salem, M.M., Abdulazim, D.O., 2017. Uric acid in the pathogenesis of metabolic, renal, and cardiovascular diseases: a review. J. Adv. Res. 8 (5), 537–548.

Sun, Y.F., Guo, D.L., 2017. Research progress in relationship between inflammatory factors and coronary heart disease. China Mod. Med. 24 (1), 12–15.

Takamura, T.A., Tsuchiya, T., Oda, M., et al., 2017. Circulating malondialdehyde-modified low-density lipoprotein (MDA-LDL) as a novel predictor of clinical outcome after endovascular therapy in patients with peripheral artery disease (PAD). Atherosclerosis 263, 192–197.

Tseng, C.J., Lu, C.J., Chang, C.C., et al., 2017. Integration of data mining classification techniques and ensemble learning to identify risk factors and diagnose ovarian cancer recurrence. Artif. Intell. Med. 78, 47–54.

Wan, J., Wang, Z., Ye, J., et al., 2017. Research progress in the role and its molecular mechanism of interleukin-6 in cardiovascular disease. Guangxi Med. J. 39 (4), 513–515.

Wang, W.Y., Ding, S.L., Song, L.H., et al., 2016. Application of neural network and support vector machine in cognitive diagnosis. Psychol. Sci. 39 (04), 777–782.

Wang, B.S., Li, D.Y., Chen, X.G., et al., 2017. Application of D-dimer, fibrinogen, etc. joint detection in diagnosis of acute myocardial infarction. Chin. J. Lab. Diagnosis 21 (2), 205–207.

Wildenberg, M.E., Koelink, P.J., Diederen, K., et al., 2017. The ATG16L1 risk allele associated with Crohn's disease results in a Rac1-dependent defect in dendritic cell migration that is corrected by thiopurines. Mucosal Immunol. 10 (2), 352–360.

Xiao, J.C., Luo, J., Zhang, R.S., et al., 2016. Clinical application of PAPP-A and hs-CRP in acute coronary syndrome. China J. Pharmacy-Clin. Collect. 36, 262.

Xu, M.X., Liu, C., He, Y.M., et al., 2017. Long-term statin therapy could be efficacious in reducing the lipoprotein (a) levels in patients with coronary artery disease modified by some traditional risk factors. J. Thoracic Dis. 9 (5), 1322–1332.

Yang, S.H., Du, Y., Zhang, Y., et al., 2017. Serum fibrinogen and cardiovascular events in Chinese patients with type 2 diabetes and stable coronary artery disease: a prospective observational study. BMJ Open 7 (6), e015041.

Yang, Y., Lian, S.J., Zhang, B.H., et al., 2017. Study on the correlation between serum lipoprotein (a) concentration and carotid atherosclerotic plaque in patients with hypertension. J. Navy Med. 38 (3), 229–231.

Yuan, B., Zhang, L.J., Li, Q., et al., 2016. Correlation between serum amyloid A1 gene polymorphism and carotid atherosclerosis in patients with atherosclerotic thrombotic cerebral infarction. Chin. J. Brain Dis. Rehabilitation 6 (02), 65–68.

Zeng, Q.F., Lu, H., 2016. Relationship between carotid atherosclerotic plaque of acute cerebral infarction patients and IL-6 and CRP. Chin. J. Pract. Nervous Dis. 19 (7), 17–19.

Zhao, C., Chen, X.D., Zhang, J.C., et al., 2017. Detection of coronary artery lesions based on one- class support vector machines. Chin. J. Lasers 44 (05), 168–175.

Zhou, J., Zhou, Y., Cao, S., et al., 2016. Multivariate logistic regression analysis of postoperative complications and risk model establishment of gastrectomy for gastric cancer: a single-center cohort report. Scand. J. Gastroenterol. 51 (1), 8–15.

Zong, Y.Y., Xu, H., Xu, W., et al., 2015. Logistic regression and ROC curve analysis of the value of serum DKK1, GP73 and AFP in the diagnosis of primary liver cancer. Lab. Med. 6, 559–563.