Acute coronary syndrome screening in patients presenting with arteriosclerosis in health check-ups: a case-control study

Xiaoxing Li, Fangkun Yan, Xinhui Liu, Mingzhuo Li, Jiangbing Li, Yugo Chen, Chuanbao Li

ABSTRACT

Objectives This research aimed to develop a simple and effective acute coronary syndrome (ACS) screening model in order to intervene early and focus on prevention in patients presenting with arteriosclerosis.

Design A case–control study.

Setting The study used a cross-sectional survey to collect data from 2243 patients who completed anonymous electronic medical record (EMR) data and coronary angiography was gathered at a hospital in Shandong Province between December 2013 and April 2016.

Participants Adults 18 years old and above diagnosed as ACS or non-ACS according to the records in hospital EMR database, and with completed basic information (age and sex).

Predictors 54 laboratory biomarkers and demographic factors (age and sex).

Statistical analysis A dataset without missing data of all patients’ laboratory indicators and demographic factors was divided into training set and validation set after being balanced. After the training set balanced, area under the curve of random forest (AUCRF) and least absolute shrinkage and selection operator (LASSO) regression were used for feature extraction. Then two set random forest models were established with the different feature sets, and the process of comparison and analysis was made to evaluate models for the optimal model including sensitivity, accuracy and AUC receiver operating characteristic curves with the internal validation set.

Main outcome measures To establish an ACS screening model.

Results An RF model with 31 features selected by LASSO with an AUC of 0.616 (95% CI 0.685 to 0.777), a sensitivity of 0.832 and an accuracy of 0.714 in the validation set. The other RF model with 27 features selected by AUCRF with an AUC of 0.621 (95% CI 0.651 to 0.797), a sensitivity of 0.849 and an accuracy of 0.728 in the validation set.

Conclusions The established ACS screening model with 27 clinical features provides a better performance for practical solution in predicting ACS.

INTRODUCTION

Acute coronary syndrome (ACS) is one of the leading causes of cardiovascular disease-related death and is encountered frequently in the emergency department (ED). It is a comprehensive disease mainly caused by disruption of coronary artery plaque and consequent thrombosis-induced severe coronary artery stenosis or occlusion. Despite major improvement in medical treatment, the in-hospital diagnosis of ACS for patients complaining of chest pain is time-consuming and expensive. The potential risk of some procedures, such as invasive coronary angiography (CAG), may outweigh the diagnostic benefits. Furthermore, some major adverse cardiac events, including ACS, do not present with premorbid clinical manifestations. In such a case, the opportunity for early diagnosis may be lost for those patients without obvious clinical symptoms, leading to poor prognosis. Thus, a case-control study for developing simple and accurate screening methods for ACS performed in regular health check-ups was essential for patients with arteriosclerosis (AS), which can prompt the treatment and avoid the risk of fatal complications as well as improve the prognosis of patients.

Although several risk factors are well known to be associated with ACS, such as the body mass index, smoking status, alcohol consumption, hypertension and diabetes, but because of the authenticity of...
Open access

Li X, et al. BMJ Open 2022;12:e062596. doi:10.1136/bmjopen-2022-062596

Figure 1A  A flow chart for study population selection. AUCRF, area under the curve of random forest; EMR, electronic medical record; LASSO, least absolute shrinkage and selection operator; RF, random forest.

data on these risk factors and the role of long-term effects cannot be quantified, such as cigarette consumption in terms of the number of cigarettes per day and duration of smoking as well as alcohol, leading these risk factors may not have high sensitivity and specificity for ACS screening individually or in some combinations. Many biomarkers in clinical practice are preferred in the ACS diagnostic process. In the early years, several biomarkers for ACS detection have been proposed, such as high-sensitivity cardiac troponin (hs-cTn). With high sensitivities and specificities, hs-cTn has been the criterion biomarkers for acute myocardial infarction (MI). Because elevated hs-cTn indicates myocardial injury, patients with elevated hs-cTn usually first go to hospital to receive treatment.
as early as possible. Hs-cTn makes little effort in people having regular medical examinations who are usually no myocardial injury leading to a normal outcome. Therefore, the selection of an optimal subset of common biomarkers in clinical practice is an important step for the construction of the ACS screening model.

Random forest (RF) is a widely used machine learning method for data classification, and has been applied in clinical settings to predict disease and shown to have higher accuracy for diagnosis than some classical methods, such as logistic regression.

The least absolute shrinkage and selection operator (LASSO) regression model was designed to overcome the shortage of multiple regression analysis in high-dimensional data leading to predictions with large variance.

CAG is the gold standard in the diagnosis of coronary stenosis. However, CAG is an invasive test with strict indications that render it difficult to use as a population screening tool. Furthermore, it is difficult to determine the nature of the plaque on CAG. In contrast, coronary CT angiography (CCTA) is a non-invasive examination that can also detect coronary stenosis. Abdulla et al compared CCTA with CAG and concluded that CCTA has high accuracy in Coronary Heart Disease (CHD) screening and can, to some degree, safely substitute CTA as a screening tool for ACS. The latter studies demonstrated the accuracy of CCTA in detecting coronary stenosis, thereby laying the foundation for the use of CCTA as an ACS-screening tool.

In this study, we proposed a simple and accurate model using RF to screen ACS for health check-ups to have intervene early and focus on prevention. The flow chart of the

**Table 1** Characteristics of the baseline population between ACS and non-ACS patients

| Variables (Group) | Non-ACS | ACS | P value |
|------------------|---------|-----|---------|
| n, (%)           | 452 (20.2) | 1791 (79.8) | 0.393 |
| sex, %           |         |     |         |
| female           | 155 (34.3) | 574 (32.0) | 0.191 |
| male             | 297 (65.7) | 1217 (68.0) | 0.051 |
| Age, years       | 62.79 (10.57) | 61.74 (10.16) | 0.177 |
| AG_R, %          | 1.69 (0.28) | 1.67 (0.29) | 0.604 |
| ALB, g/L         | 42.54 (3.43) | 42.26 (3.90) | 0.767 |
| WCC, 10^9/L      | 6.44 (1.65) | 6.39 (1.74) | 0.056 |
| MON_R, %         | 6.45 (1.69) | 6.48 (1.93) | 0.087 |
| MON_C, 10^9/L    | 0.41 (0.15) | 0.41 (0.15) | 0.026 |
| LDL_C, mmol/L    | 2.62 (0.86) | 2.61 (0.86) | 0.001 |
| Ca, mmol/L       | 2.30 (0.12) | 2.30 (0.11) | 0.015 |
| TG, mmol/L       | 1.51 (0.91) | 1.59 (1.09) | 0.017 |
| HDL_C, mmol/L    | 1.19 (0.27) | 1.18 (0.25) | 0.003 |
| ALT, U/L         | 27.17 (35.06) | 33.97 (37.16) | <0.001 |
| AST, U/L         | 23.50 (16.16) | 27.83 (25.70) | 0.051 |
| RDW, %           | 12.96 (0.85) | 12.87 (0.85) | 0.051 |
| HCT, %           | 11.68 (18.28) | 9.60 (17.09) | 0.023 |
| APTT_R           | 1.08 (0.15) | 1.09 (0.16) | 0.249 |
| APTT_S           | 31.85 (8.14) | 31.83 (6.42) | 0.968 |
| CR, μmol/L       | 71.30 (17.09) | 70.87 (23.89) | 0.72 |
| CK, U/L          | 84.10 (57.30) | 87.73 (139.45) | 0.588 |
| K, mmol/L        | 4.10 (0.39) | 4.15 (0.38) | 0.009 |
| IBIL, μmol/L     | 8.40 (3.94) | 8.33 (4.13) | 0.739 |
| AKP, U/L         | 69.85 (20.42) | 73.35 (22.57) | 0.003 |
| LYM_R, %         | 29.25 (8.02) | 28.83 (8.35) | 0.335 |
| LYM_C, 10^9/L    | 1.84 (0.59) | 1.80 (0.63) | 0.169 |
| Mg, mmol/L       | 0.88 (0.08) | 0.88 (0.08) | 0.686 |
| BUN, mmol/L      | 5.17 (1.53) | 5.03 (1.59) | 0.095 |
| UA, μmol/L       | 332.50 (96.06) | 318.88 (84.64) | 0.003 |
| TT_S, second     | 15.33 (10.68) | 15.32 (9.84) | 0.982 |
| TT_R, %          | 0.97 (0.09) | 0.99 (0.48) | 0.281 |
| PT_INR, %        | 1.01 (0.23) | 0.99 (0.09) | 0.02 |
| PT_R, %          | 1.01 (0.33) | 0.99 (0.09) | 0.012 |
| PT_ACTIVITY, %   | 99.06 (15.17) | 99.57 (13.92) | 0.491 |
| MCV, fL          | 91.47 (4.65) | 91.58 (4.51) | 0.654 |
| MCH, pg          | 31.32 (1.96) | 31.35 (1.95) | 0.758 |
| MCHC, g/L        | 342.39 (12.03) | 342.34 (12.96) | 0.94 |
| MPV, fL          | 9.67 (1.29) | 9.58 (1.23) | 0.2 |
| GLU, mmol/L      | 5.64 (1.60) | 5.60 (1.84) | 0.629 |
| GLB, mmol/L      | 25.74 (3.82) | 25.85 (3.88) | 0.581 |
| BAS_C, 10^9/L    | 0.02 (0.02) | 0.02 (0.02) | 0.702 |
| EOS_C, 10^9/L    | 0.16 (0.15) | 0.15 (0.14) | 0.172 |
| HGB, 10^9/L      | 138.78 (14.66) | 139.37 (15.37) | 0.458 |

Continued
study population selection can be seen in figure 1. And, a case example shows in figure 2.

**METHODS**

**Patient and public involvement**

No patient or public involved.

**Data source and participants**

It was a cross-sectional, observational, single-centre study that all patients who were adults 18 years old and above diagnosed as ACS or non-ACS according to the records in hospital electronic medical record (EMR) database, and with completed basic information (age and sex) from December 2013 to April 2016. Patients with known connective tissue disorders, pulmonary embolism, dissection, systemic inflammatory conditions, prior acute MI or prior heart surgery were excluded. Patients with AS were labelled by CCTA or CAG with any part stenosis above 50%. They were all accepted by at least one of CCTA or CAG.

**INCLUSION CRITERIA**

(1) Definitely diagnosed as AS according to the records of CAG in hospital EMR database whose coronary arteries were more than 50% stenosis; (2) at least 18 years old; (3) without missing basic information (age and sex) and (4) availability of medical records of clinical examinations (missing rate below 50%).

**EXCLUSION CRITERIA**

(1) Age <18 years; (2) hs-cTnI elevation when the first admitted hospital blood test; (3) previous history of MI, CAG or Percutaneous Transluminal Coronary Intervention (PCI); (4) Cases with missing information of seven

| Variables (Group) | Non-ACS | ACS | P value |
|-------------------|---------|-----|---------|
| PLT, 10^9/L       | 221.30 (61.50) | 223.27 (62.28) | 0.547 |
| PDW, fl           | 15.00 (2.06) | 15.05 (2.05) | 0.634 |
| DBIL, μmol/L      | 3.97 (1.57) | 3.90 (1.46) | 0.377 |
| NEU_R, %          | 61.33 (8.79) | 61.86 (8.93) | 0.26  |
| NEU_C, 10^9/L     | 4.00 (1.38) | 4.01 (1.46) | 0.88  |
| TC, mmol/L        | 4.29 (1.08) | 4.28 (1.10) | 0.799 |
| TBL, umol/L       | 12.37 (5.31) | 12.23 (5.33) | 0.61  |
| TBA, μmol/L       | 6.79 (5.19) | 6.26 (5.00) | 0.049 |
| TP, g/L           | 68.23 (5.12) | 68.02 (5.42) | 0.451 |
| DD, μg/ml         | 0.39 (1.97) | 0.22 (0.76) | 0.004 |
| RCC, x10^12/L     | 4.44 (0.48) | 4.45 (0.49) | 0.667 |
| hs_CTNI, ng/mL    | 0.02 (0.01) | 0.02 (0.01) | 0.011 |
| PT_S, second      | 10.78 (3.60) | 10.54 (0.84) | 0.008 |
| CKMB, ng/mL       | 1.44 (1.11) | 1.96 (8.36) | 0.183 |
| PA, mg/dl         | 23.03 (5.18) | 22.46 (4.87) | 0.029 |

Data are presented as means±SDs, medians (IQR) or percentage.

ACS, acute coronary syndrome; AG_R, albumin/globulin ration; AKP, alkaline phosphatase; ALT, alanine aminotransferase; APTT, activated partial thromboplastin time; APTT-IR, APTT ratio; AST, aspartate amino transferase; BAS_C, basophilic granulocyte count; BAS_R, basophilic granulocyte ratio; BUN, blood urea nitrogen; Ca, calcium ion; CK, creatine kinase; CKMB, creatine kinase isoenzymes; CR, creatinine; DBIL, direct bilirubin; EOS_C, eosinophil count; EOS_R, eosinophil ratio; GLB, globulin; GLU, glucose; HCT, haematocrit; HDL, high-density lipoprotein cholesterol; HGB, haemoglobin; hs_CTNI, high-sensitivity cardiac troponin; IBL, indirect bilirubin; K, serum kalium; LDL_C, low-density lipoprotein cholesterol; LYM_C, lymphocyte count; LYM_R, lymphocyte ratio; MCH, mean corpuscular haemoglobin; MCHC, mean corpuscular haemoglobin concentration; MCV, mean corpuscular volume; Mg, magnesium ion; MON_C, monocyes count; MON_R, Monocytes ratio; MPV, mean platelet volume; NEU_C, neutrophil count; NEU_R, neutrophil ratio; PA, prealbumin; PDW, mean platelet distributing width; PLT, platelet; PT_ACTIVITY, prothrombin time activity; PTINR, PT international standard ratio; PT_R, PT ratio; PT_S, prothrombin time; RCC, red cell count; RDW, red cell distribution width; TBA, total bile acid; TBL, total bile acid; TC, total cholesterol; TG, triglyceride; TP, total protein; TT_S, thrombin time ratio; TT_S, thrombin time; UA, Uric acid; WCC, white cell count.

**Figure 3** Dimensionality reduction and feature selection of variables. AUC, area under the curve.

**Figure 4** Feature selection using AUCRF. AUC, area under the curve; AUCRF, AUC of random forest; OOB, out-of-bag.
Outcomes

Our outcome was ACS. According to the consensus guidelines of the European Society of Cardiology, patients with a diagnosis of unstable angina pectoris and acute MI, including ST-segment elevation MI and non-ST-segment elevation MI, were considered as having ACS. All patients completed CAG were identified by a professional cardiovascular panel during the hospitalisation period.

Predictors

A number of 54 laboratory biomarkers and demographic factors (age and sex) were involved according to the risk factor playing a role in the pathological way of ACS. All laboratory tests were performed under the supervision of a professional cardiovascular panel during hospitalisation. For those laboratory biomarkers that were evaluated more than once, the first measurement after admission was selected in order to eliminate the impact of medical treatments during the hospitalisation. (The time elapsed from biomarkers evaluation and ACS or atherosclerosis detection: 21.35±36.24 hours).

Since all the predictors were objective values, and the outcomes were judged after the measurement of the predictors, it can ensure the assessment of predictors for the outcome and other predictors was blind.

Statistical analysis

Data preprocessing

Original data (n=2243) were retrospectively collected for patients with the hs-cTn value under the referent upper limit undergoing CAG or CCTA according to the inclusion and exclusion criteria from December 2013 to April 2016. The baseline population data were selected from the original data. Temporary data were formed after deleting columns with missing values greater than 20% of each column and rows with missing values greater than 20% of each row. By using the technique of multiple imputation to handle the missing values in the temporary data, the constructed baseline population data are shown in table 1.

The data of baseline population was divided into training sets and validation sets in a ratio of 9:1. Imbalanced data usually lead to overly optimistic performance estimates in machine learning like RF, KNN, deep learning. Model performance metrics are also involved with bias that positive predictive values remain unwatchable low while high area under the curve (AUC) achieved by models. We balanced within the training set only using the package of ROSE. The Validation set remains the original imbalanced class in order to real reflect the true anticipated.

We made a comparison between the group with and without ACS; in particular, the continuous biomarkers were compared using t-tests (for normally distributed variables) or the Wilcoxon rank sum test (for variables with skewed distributions) after testing for data normality with the Kolmogorov-Smirnov test, whereas the categorical

| Table 2 | The importance of different model |
|--------|----------------------------------|
| Variables | Weight of LASSO | MDG of RF |
| ALT | 0.002953 | 19.93027 |
| AST | 0.005179 | 16.43825 |
| PLT | −0.00039 | 16.31935 |
| CK | −0.00015 | 16.31204 |
| GLU | −0.05461 | 15.35616 |
| MCV | 0.016542 | 15.02556 |
| TBA | −0.01551 | 15.17738 |
| MPV | −0.05141 | 15.17738 |
| K | 0.347483 | 14.84482 |
| TG | 0.199565 | 14.83493 |
| TC | −0.05141 | 14.69868 |
| AKP | 0.001516 | 14.47648 |
| UA | −0.00119 | 14.08369 |
| LDL_C | −0.04357 | 13.58772 |
| Cr | . | 13.58071 |
| PA | −0.03436 | 13.33002 |
| LYM_R | . | 13.17961 |
| MON_C | . | 13.14967 |
| LYM_C | . | 13.11114 |
| HCT | −0.00662 | 13.0431 |
| MOHC | −0.00336 | 12.98987 |
| HDL_C | . | 12.86844 |
| RCC | . | 12.78484 |
| MON_R | 0.004525 | 12.77796 |
| age | −0.00152 | 12.72252 |
| ALB | −0.00371 | 12.69232 |
| BAS_C | −1.78757 | |
| Ca | 0.153915 | |
| CKMB | 0.011006 | |
| DBIL | 0.032123 | |
| DD_I | −0.06767 | |
| EOS_C | −0.30583 | |
| hs_CTNI | 0.63572 | |
| NEU_C | 0.005345 | |
| PDW | −0.01682 | |
| PT_R | −1.17256 | |
| Sex | 0.225525 | |
| TT_R | 0.08577 | |

AKP, alkaline phosphatase; ALT, alanine aminotransferase; AST, aspartate aminotransferase; BAS_C, basophilic granulocyte count; BUN, blood urea nitrogen; CK, creatine kinase; CKMB, creatine kinase isoenzymes; Cr, creatinine; DBIL, direct bilirubin; EOS_C, eosinophil count; HCT, haematocrit; HDL_C, high-density lipoprotein cholesterol; LASSO, least absolute shrinkage and selection operator; LDL, low-density lipoprotein cholesterol; LYM_C, lymphocyte count; LYM_R, lymphocyte ratio; MCHC, mean corpuscular haemoglobin concentration; MCV, mean corpuscular volume; MDG, mean decrease in gini; MON_C, monocytes count; MON_R, monocytes ratio; NEU_C, neutrophil count; PA, prealbumin; PDW, platelet distributing width; PLT, platelets; RCC, red cell count; RF, risk factor; TBA, total bile acid; TC, total cholesterol; TT_R, thrombin time ratio; UA, uric acid.
variables were assessed using $\chi^2$ test. A two-tailed $p<0.05$ was considered statistically significant.

**Feature selection**

In order to construct a more robust and automated model with the most relevant features. Two different feature selection techniques, including LASSO, and AUC of RF (AUCRF) were used to determine the optimal subset of characters.

AUCRF was an algorithm using optimising the AUCRF instead of the classification error for accuracy of RF. AUC-RF builds an RF with all the features, ranking importance of features by the Gini index, then backwards elimination. The AUC-RF algorithm gives out both the out-of-bag (OOB) error and the OOB-AUC based on the OOB predictions.

LASSO regression algorithm provides a method dealing with high-dimensional data overcoming large variance caused by multiple regression analysis. LASSO penalises the coefficients by setting as many coefficients as possible to zero to extract as few features as possible.

**RF models**

RF, first introduced by Breiman, was a collection of classification and regression trees. This methodology is used to address two main classes of problems: the assessment and rank variables with respect to their ability to predict the response and the construction of a prediction rule for a supervised learning problem. RF variable importance measures can successfully identify predictors involved in interactions. In addition, it overcomes the overfitting problem of individual decision tree and consistently offers the highest prediction accuracy in the setting of classification even with highly correlated variables compared with other models.

Two set models were developed with the methods of RF using features selected by LASSO and AUCRF. Lasso. The optimal parameters of RF were set by mtry=$\sqrt{\text{number of variables}}$ and ntree was large enough to discriminate.

**Models evaluation**

We compared the performances of two models with AUC, accuracy, sensitivity, specificity and F1-Score.

The statistical analyses in this study were performed using R software V.4.1.1 with packages caret, randomForest, mice, tableone, rose and receiver operating characteristic (pROC). We used the TRIPOD checklist when writing our report.

**RESULTS**

The demographic and clinical characteristics of the patients in the whole dataset are shown in Table 1. Among total 2243 cases, 1791 (14.3%) patients had ACS. Thirteen laboratory indicators showed significant differences between the case (ACS) and control (non-ACS) groups. In terms of sex, males accounted for 65.7% and females accounted for 34.3% of patients with non-ACS, while males accounted for 68% and females accounted for 32% of patients with ACS. The result showed statistically significant differences in aspartate aminotransferase (AST), alanine aminotransferase (ALT), alkaline phosphatase (AKP), uric acid (UA), DD-I, prothrombin time, K, hs-CTNI, prothrombin time ratio, PT_international standard ratio, haematocrit (HCT), prealbumin (PA), total bile acid (TBA), but the other 43 variables showed no significant differences.

In the feature selection process of LASSO, the best lambda was obtained shown in Figure 3 and the lambda=−4.95 was chosen to solve the multicollinearity problem. The result of feature selection with AUCRF was shown in Figure 4. Twenty-seven variables were selected with the highest OOB-AUC of 0.9544. We listed all the selected variables in Table 2. In the LASSO feature selection term, 31 variables had a weight of none zero. In the AUCRF feature selection term, all the variables ranked according to the mean decrease in gini.

To evaluate the prediction performance of the two models, a series of indicators including AUC, 95% CI, accuracy, sensitivity, specificity, F1-score were listed in Table 3. Model B had a better performance. A ROC figure also provided to evaluated (figure 5).

Finally, model B constructed variable selected by AUCRF had the better prediction performance.

**DISCUSSION**

Our study shows that 19 variables appear simultaneously in the two reconstructed models including ALB, age, monocytes ratio, mean corpuscular haemoglobin concentration (MCHC), HCT, PA, LDL_C, UA, AKP, TG, K, blood urea nitrogen (BUN), TBA, mean corpuscular volume, GLU, CK, platelets (PLT), AST, ALT, while age is an independent predictor for ACS and diabetes was known as a risk factor for both ACS and AS. Although ACS is associated with sudden clinical manifestations, disease progression may occur gradually in a clinically asymptomatic manner over many years; in fact, the

| Model | AUC   | 95% CI       | Accuracy | Sensitivity | Specificity | F1-Score | Cut-off |
|-------|-------|--------------|----------|-------------|-------------|----------|---------|
| A     | 0.616 | (0.650 to 0.772) | 0.714    | 0.832       | 0.244       | 0.823    | 0.617   |
| B     | 0.621 | (0.664 to 0.785) | 0.728    | 0.849       | 0.244       | 0.833    | 0.622   |

Model A: constructed with feature selected by LASSO; Model B: constructed with feature selected by AUCRF.

AUC, the area under the curve; AUCRF, AUC of random forest; LASSO, least absolute shrinkage and selection operator.
rupture of plaques is usually asymptomatic. Inflammation plays an important role in plaque instability as well as the pathogenesis of ACS. With vascular endothelial injury, white blood cells in the peripheral circulation, including neutrophil, eosinophil and monocytes are guided to the site of lesion to promote the formation of plaques. Mature atherosclerotic plaques contain large amounts of lipids, foam cells which are macrophages or smooth muscle cells with large amounts of fat, as well as proliferating smooth muscle cells and matrix components (collagen, elastin). The further development of the formed plaque shows the enlargement of the lipid nucleus and the thinning of the vascular cap. Whenever plaque ruptures the PLT adhesion and aggregation further occurs, and thrombosis is formed. PLT are involved in the production of blood clots and provide mediators for the formation and maintenance of local inflammatory responses. Since PLT play a certain role in systemic inflammatory response. Nozawa et al showed monocytes in circulating play an important role in the progression of coronary plaque in AMI, and suggested that the peak monocyte count might be a predictor for plaque. LDL and TG also play an important role in ACS prediction. PLT are involved in the inflammatory process, platelet distributing width (PDW) is an important and simple marker of significantly elevated PLT activation in inflammatory processes. UA has been identified as a significant determinant of ACS, Heart Failure. Ren et al suggested serum AKP may be a potential predictive biomarker. Adam et al found that K along with BUN and creatinine (Cr) has a strong relationship in diagnosis of ACS, especially in mortality-risk assessment of ACS patients. High MCHC and MPV levels may be beneficial for ACS. ALT and AST are widely present in the myocardium, brain and other organs, and related to ACS. However, the effect of TBA on ACS has not been found.

Our study shows that the other 20 variables appear only in one model, red blood cell, high-density lipoprotein cholesterol, lymphocyte count, monocytes count, lymphocyte ratio, Cr, total cholesterol, MPV in Model A and basophilic granulocyte count, PT_R, eosinophil count, DD_I, PDW, neutrophil count, creatine kinase isoenzymes, direct bilirubin, thrombin time ratio, Ca, sex, hs_CTNI in model B. But they still play a different role in the development of MI.

Biomarkers in clinical practice are usually in complex interaction structures or are highly correlated. In this study, we used AUCRF and LASSO to assess the importance of each variable and constructed the screening model, which can successfully identify predictors involved in interactions and consistently offers the highest prediction accuracy in the setting of classification even with highly correlated variables compared with other models. And RF is more suitable for feature selection.

Our study shows that through optimising glycogen levels, blood pressure control, lipid levels and ensuring those variables used for constructing models can be a better secondary prevention, and reduce major adverse cardiovascular events in the future.

Our study has several limitations. First, model B with a better prediction performance only has an AUC of 0.621, might be caused by the small amount of data, unbalanced data distribution and other reasons. ECG has always played a vital role in diagnosis of ACS. It is necessary to combine ECG to help make a full judgement when patients with AS have a regular health check at this stage. It is reasonable to believe that the effectiveness of our ACS-screening model can be further improved by introducing new variables, a more complete dataset and a greater number of indices in future studies. Second, only indicators with a missing data rate of less than 20% were included in this study. Therefore, indicators with high missing data rates and high discriminant values could have been missed, and these can be explored in future studies. Third, datasets included in this study involved hospitalised AS patients who received CAG during hospitalisation, and
this population may not be fully representative of all AS patients, a further multicentre study should be done. Finally, the RF algorithm, while robust, has some inherent defects, such as the difficulty in the explanation of the contribution of a single predictor.

CONCLUSION
The established ACS screening model with 27 clinical features provides a better performance for practical solution in predicting ACS, it can intergrate in laboratory test equipment for diagnosis-early-warmin and focus on prevention in regular health check-ups.

Author affiliations
1 Department of Geriatrics, Qilu Hospital of Shandong University, Jinan, Shandong, China
2 Department of Emergency and Chest Pain Center, Qilu Hospital of Shandong University, Jinan, Shandong, China
3 Department of Biostatistics, School of Public Health, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China
4 Institute for Medical Dataology, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China
5 Center for Big Data Research in Health and Medicine, Shandong Qianfoshan Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China
6 Department of Cardiology, Shandong Provincial Hospital, Cheeloo College of Medicine, Shandong University, Jinan, Shandong, China

Acknowledgements
This study is a joint effort of many investigators and staff members, and their contribution is gratefully acknowledged. We especially thank the Health Management Center, as well as the individuals who participated in this study. We are also grateful for the support of the Qila National Key Research and Development Program of China, the Shandong Provincial Natural Science Foundation of China, Shandong Provincial Key Research and Development project, the National Natural Science Foundation of China, the Taishan Pandeng Scholar Program of Shandong Province, and the National Natural Science Foundation of Shandong Province.

Contributors
FY and XL analysed data. XL and FY wrote the manuscript. ML helped to collect data. JL made efforts on supervision for data quality. YC and CL designed the study, assisted with the methods and data analysis discussion, and reviewed all drafts. CL is the guarantor of the article taking full responsibility for the finished work and the conduct of the study, had access to the data, and controlled the decision to publish.

Funding
This work was supported by the National Natural Science Foundation of China (Grant number 82070388 and 82170442), the Taishan Pandeng Scholar Program of Shandong Province (Grant number tsps20181220) and the National Natural Science Foundation of Shandong Province (Grant number ZR2020MH035).

Competing interests
None declared.

Patient and public involvement
Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication
Consent obtained directly from patient(s).

Ethics approval
This study involves human participants and was approved by the Public Health Ethics Committee of Shandong University (Approval No. 20180801). Participants gave informed consent to participate in the study before taking part.

Provenance and peer review
Not commissioned; externally peer reviewed.

Data availability statement
Data are available on reasonable request. Data can be made available to researchers approved at request of the corresponding author, with a signed data access agreement.

Open access
This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: http://creativecommons.org/licenses/by-nc/4.0/.

REFERENCES
1 Benjamin EJ, Muntner P, Alonso A, et al. Heart disease and stroke Statistics-2019 update: a report from the American Heart association. Circulation 2019;139:e66–28.
2 Kimura K, Kimura T, Ishihara M, et al. JCS 2018 guideline on diagnosis and treatment of acute coronary syndrome. Circ J 2019;83:1085–196.
3 Apple FS, Wu AHB, Mair J, et al. Future biomarkers for detection of ischemia and risk stratification in acute coronary syndrome. Clin Chem 2005;51:810–24.
4 Eigenday Y, Mahmoud AN, Mansoor H, et al. Early invasive versus initial conservative strategies for women with non-ST-elevation acute coronary syndromes: a nationwide analysis. Am J Med 2017;130:1059–67.
5 Mirza AJ, Taha AF, Khdir BR. Risk factors for acute coronary syndrome in patients below the age of 40 years. Egypt Heart J 2018;70:233–5.
6 Ahmed MH, Awadalla H, Elmadhoun WM, et al. Prevalence and risk factors for acute coronary syndrome among Sudanese individuals with diabetes: a population-based study. Cardiol Res 2017;8:184–9.
7 Sakhi H, SMUMMM J. Admission plasma glucose level among non-diabetic patients sustaining acute coronary syndrome. Mustansirija Medical Journal 2011;10:1–7.
8 Alpert JS, Thyesen K, Antman E, et al. Myocardial infarction redefined--a consensus document of the The Joint European Society of Cardiology/ American College of Cardiology Committee for the redefinition of myocardial infarction. J Am Coll Cardiol 2000;36:959–69.
9 Lin JC, Apple FS, Murakami MM, et al. Rates of positive cardiac troponin I and creatinine kinase MB mass among patients hospitalized for suspected acute coronary syndromes. Clin Chem 2004;50:333–8.
10 Liaw A, Wiener M. Classification and regression by randomForest. R news 2002;2:18–22.
11 Amini P, Maroufizadeh S, Samani RO. Factors associated with macrosomia among singleton live-birth: a comparison between logistic regression random forest and artificial neural network methods 2016;4.
12 RJfotRSSSSB T. Regression shrinkage and selection via the LASSO 1996e;58:267–88.
13 Casella G, Ghosh M, Gill J, et al. Penalized regression, standard errors, and Bayesian lassos. Bayesian Anal 2010;5:369–411.
14 Levine GN, Bates ER, Blankenship JC. 2015 ACC/AHA/SCAI focused update on primary percutaneous coronary intervention for patients with ST-elevation myocardial infarction: an update of the 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention and the 2013 ACCF/AHA guideline for the management of ST-elevation myocardial infarction. J Am Coll Cardiol 2016;67:1235–50.
15 Wright RS, Monnahan RL, Kopecky SL, et al. Cardiac catheterization reduces resource utilization in patients with chronic chest pain. Circ Cardiovass Interv 2000;4:365–8.
16 Achenbach S. Computed tomography coronary angiography. J Am Coll Cardiol 2006;48:1919–28.
17 Meljiboom WB, Meij MFL, Schuijt JD, et al. Diagnostic accuracy of 64-slice computed tomography coronary angiography: a prospective, multicenter, multivendor study. J Am Coll Cardiol 2008;52:2153–44.
18 Abdulla J, Abildstrom SZ, Gotzsche O, et al. 64-multislice detector computed tomography coronary angiography as potential alternative to conventional coronary angiography: a systematic review and meta-analysis. Eur Heart J 2007;28:3042–50.
19 Roofi M, Patrono C, Collet J-P, et al. 2015 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation: Task force for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation of the European Society of cardiology (ESC). Eur Heart J 2016;37:267–315.
20 Zhang Z. Multiple imputation with multivariate imputation by chained equation (MICE) package. Ann Transl Med 2016;4:30.
21 Lunardon N, Menardi G, NURJ T, Rose: a package for binary imbalanced learning 2014;6.
22 Wang SB. Machine learning to advance the prediction, prevention and treatment of eating disorders. Eur Eat Disord Rev 2021;29:683–91.
23 Breiman L. Random forests. Mach Learn 2001;45:5–32.
24 Boulesteix A-L, Janitza S, Kruppa J, et al. Overview of random forest methodology and practical guidance with emphasis on prediction.
computational biology and bioinformatics. *WIREs Data Mining Knowl Discov* 2012;2:493–507.
25. Speiser JL, Miller ME, Tooze J, et al. A comparison of random forest variable selection methods for classification prediction modeling. *Expert Syst Appl* 2019;134:93–101.
26. Fernández-Delgado M, Cernadas E, Barro S. Do we need hundreds of classifiers to solve real world classification problems? *Journal of Machine Learning Research* 2014;15:3133–81.
27. 2016 7th IEEE international conference on software engineering and service science (ICSESS). Variable selection using mean decrease accuracy and mean decrease gini based on random forest. IEEE, 2016.
28. Kuhn M. Caret: classification and regression training. *Astrophysics Source Code Library* 2015.
29. van Buuren S, Boshuizen HC, Knook DL. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat Med* 1999;18:681–94.
30. Flury BK, HHTAS R. Standard distance in univariate and multivariate analysis. *JSTOR* 1986;40:249–51.
31. Robin X, Turck N, Hainard A, et al. pROC: an open-source package for R and S+ to analyze and compare ROC curves. *BMC Bioinformatics* 2011;12:1–8.
32. Collins GS, Reitsma J, Altman DG, et al. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. *Br J Surg* 2015;102:148–58.
33. Rich MW, Bosner MS, Chung MK, et al. Is age an independent predictor of early and late mortality in patients with acute myocardial infarction? *Am J Med* 1992;92:7–13.
34. Newby UK, Jesse RL, Babb JD, et al. ACCF 2012 expert consensus document on practical clinical considerations in the interpretation of troponin elevations: a report of the American College of cardiology Foundation Task force on clinical expert consensus documents. *J Am Coll Cardiol* 2012;60:2427–63.
35. Nozawa N, Hibi K, Endo M, et al. Association between circulating monocytes and coronary plaque progression in patients with acute myocardial infarction. *Circ J* 2010;74:1384–91.
36. Budzianowski J, Pieszko K, Burchardt P, et al. The role of hematological indices in patients with acute coronary syndrome. *Dis Markers* 2017;2017:1–9.
37. Maloberti A, Biolcati M, Ruzzententi G, et al. The role of uric acid in acute and chronic coronary syndromes. *J Clin Med* 2021;10. doi:10.3390/jcm10204750. [Epub ahead of print: 16 10 2021].
38. Ren Y, Li X, Wang S, et al. Serum alkaline phosphatase levels are associated with coronary artery calcification patterns and plaque vulnerability. *Catheter Cardiovasc Interv* 2021;97 Suppl 2:1055–62.
39. Adam AM, Naas SAR, Merchant AZ, et al. Efficacy of serum blood urea nitrogen, creatinine and electrolytes in the diagnosis and mortality risk assessment of patients with acute coronary syndrome. *Indian Heart J* 2018;70:353–9.
40. Karaman S, Coolkun A. Do MCHC, MPV, and procalcitonin levels determine prognosis in acute coronary syndrome? *Emerg Med Int* 2019:2019:1–8.
41. Güntekin Ünal, Tosun V, Kiliç AY, et al. ST segment elevation myocardial infarction (STEMI) patients are more likely to achieve lipid-lowering treatment goals: a retrospective analysis of patients presenting with first acute coronary syndromes. *Medicine* 2018;97:e12225.
42. Breiman L, Friedman JH, Olshen RA. *Classification and regression trees*. Routledge, 2017.
43. Hilfiker-Kleiner D, Mey C, Schieffer E, et al. Recovery from postpartum cardiomyopathy in 2 patients by blocking prolactin release with bromocriptine. *J Am Coll Cardiol* 2007;50:2354–5.