Machine learning algorithm can provide assistance for the diagnosis of non-ST-segment elevation myocardial infarction

Lian Qin,1 Quan Qi,2 Ainiwaer Aikeliyaer,1 Wen Qing Hou,2 Chang Xin Zuo,2 Xiang Ma1

ABSTRACT

Introduction Our aim was to use the constructed machine learning (ML) models as auxiliary diagnostic tools to improve the diagnostic accuracy of non-ST-elevation myocardial infarction (NSTEMI).

Materials and methods A total of 2878 patients were included in this retrospective study, including 1409 patients with NSTEMI and 1469 patients with unstable angina pectoris. The clinical and biochemical characteristics of the patients were used to construct the initial attribute set. SelectKBest algorithm was used to determine the most important features. A feature engineering method was applied to create new features correlated strongly to train ML models and obtain promising results. Based on the experimental dataset, the ML models of extreme gradient boosting, support vector machine, random forest, naive Bayesian, gradient boosting machines and logistic regression were constructed. Each model was verified by test set data, and the diagnostic performance of each model was comprehensively evaluated.

Results The six ML models based on the training set all play an auxiliary role in the diagnosis of NSTEMI. Although all models taken for comparison performed differences, the extreme gradient boosting ML model performed the best in terms of accuracy rate (0.95±0.014), precision rate (0.94±0.011), recall rate (0.98±0.003) and F-1 score (0.96±0.007) in NSTEMI.

Conclusions The ML model constructed based on clinical data can be used as an auxiliary tool to improve the accuracy of NSTEMI diagnosis. According to our comprehensive evaluation, the performance of the extreme gradient boosting model was the best.

INTRODUCTION

Acute coronary syndrome (ACS) is a common cardiovascular disease associated with high complication and mortality rates,1 and the incidence of non-ST-segment elevation ACS (NSTEMI-ACS) has gradually increased in recent years.2–4 The high morbidity and mortality rate of NSTE-ACS5,6 makes it important to implement accurate diagnosis in the early stage of the disease. Despite advances in treatment of chest pain improving the hospitalisation rate and mortality of ST-segment elevation myocardial infarction (STEMI),8 the proportion of patients with NSTE-ACS hospitalised has increased. To the best of our knowledge, indistinguishable atypical features of ECG may lead to a decrease in the diagnostic accuracy of NSTEMI and unstable angina pectoris (UA).1,9 The information provided by non-specific changes in myocardial injury markers10,11 may also affect therapeutic decisions.12 Machine learning (ML), as a field in artificial intelligence (AI) research, has undergone an expansion of its applications in recent years. Multiple studies have shown that ML algorithms have high performance in cardiovascular disease risk prediction,13 imaging analysis14 and diagnosis.15 The high efficiency of the ML model in data processing makes it possible for it to be used as an auxiliary diagnostic tool for NSTEMI. Appropriate ML algorithms are expected to improve the accuracy of diagnosis and the efficiency of clinical practice, and provide information for doctors to make treatment decisions.16,17 In the present study, we constructed six ML models based on the clinical data of patients with NSTEMI and UA from two medical centres, and the diagnostic performance of each model was evaluated comprehensively. This study provides new clues for the application of ML diagnostic models in the management of NSTE-ACS.

MATERIALS AND METHODS

Participants

The clinical data of this study were derived from the chest pain centre database of the First Affiliated Hospital of Xinjiang Medical University and the First Affiliated Hospital of Shihezi University School of Medicine. The clinical data of 2878 patients diagnosed with NSTE-ACS from January 2017 to December 2019 were recorded, including 1409 patients with NSTEMI and 1469 patients with UA (figure 1A). A total of 56 clinical and laboratory features were included in the database, which were manually labelled by the Xinjiang Medical University research team using a double-blind method. The original multimodal data in the database were integrated into the verified one-dimensional structural data and the initial dataset was constructed (figure 1B).

Quality control of clinical data

The clinical data collected in this study have been reported on the China Chest Pain Center Data Reporting Platform. Data quality was ensured through design of the data collection tables, concise data definitions, central training of data entry personnel across clinical sites and remote auditing at the National Data Platform. According to the medical records of the patients, previous
medical history and family history were collected. Initial clinical presentation, standard 12-lead ECG features, myocardial enzyme profile, troponin, echocardiography and coronary angiography (CAG) were also systematically transcribed from medical records. The medical laboratory centres of the two hospitals have been certified by ISO15189 international quality system. The blood parameters were analysed by automatic blood cell analyzer (SYSMEX XN9000, Japan; UniCel DxH 800 Coulter, Beckman Coulter, USA), and the serum biochemical indexes were analysed by automatic biochemical analyzer (Roche Cobas C701, Switzerland; DxC700AU, Beckman Coulter, USA). CAG and echocardiography were performed by cardiologists with >5 years of working experience, and the results were interpreted by three cardiovascular experts to confirm the diagnosis. All data were entered in a double-blinded manner to ensure accuracy.

Inclusion and exclusion criteria of clinical data

In order to obtain the most comprehensive dataset, the clinical data included were consistent with the definition and diagnostic criteria of ‘2020 ESC Guidelines for the Management of Acute Coronary Syndromes in Patients Presenting without Persistent ST-segment Elevation’. The age range of the patients was 30–75 years, enrolling a total of 2878 participants.

(1) The clinical data of patients with aortic dissecting aneurysm, pneumothorax and other non-cardiogenic chest pain were excluded. (2) Patients with the following diseases were also excluded: liver failure, renal failure, primary tumours, severe infections and female patients during pregnancy. (3) Patients who were diagnosed with pulmonary heart disease, congenital heart disease, cardiomyopathy, severe valvular heart disease, infective endocarditis and viral myocarditis.

Preprocessing of clinical data in the initial dataset

The original dataset included 56 characteristic attributes, 30 of which were general clinical data characteristic items and the remaining 26 laboratory test results (tables 1 and 2). The data analysis package Pandas in Python was selected to read the data in the original dataset as Dataframe type, the String and Object type data in the original dataset were converted to Int and Float types that could be operated in ML operations. The outliers in the dataset were detected and filtered, and the missing values were filled by means of mode, average, median and modelling prediction. When the data result labels, medical history feature items and family history feature items were extracted, One-Hot coding was performed to convert these categorical variables into data forms that could be applied to
ML algorithms. The experimental dataset obtained after data preprocessing was used for training and testing of ML models.

Screening of clinical feature items by ML algorithms

Three types of ML algorithms were selected for the screening of clinical features, including SelectKBest, extreme gradient boosting (XGBoost) and random forest (RF). The principle of the SelectKBest algorithm is to filter out the best items from all the feature items to form a new feature dataset that can be applied to the screening of classified feature items. The XGBoost algorithm has advantages in data regression, data classification and sorting, and can also be used as an effective algorithm for data dimensionality reduction. Compared with other algorithms, XGBoost for feature selection has advantages in missing data processing. As one of the conventional algorithms, the RF algorithm has high accuracy in feature selection, avoids overfitting of feature items and has the advantages of wide applicability. Therefore, the XGB regressor and RF regressor function packages were selected to complete the screening of clinical feature items. Through comparison of the performance of three ML feature screening algorithms, the optimal algorithm was selected to construct the experimental dataset.

Construction of ML models

According to the weight of the classification contribution and the correlation coefficient, the clinical feature items in the experimental dataset were classified and sorted. The clinical data of the First Affiliated Hospital of Xinjiang Medical University were used for training and verification of ML models, and the data of the First Affiliated Hospital of Shihezi University School of Medicine were used to test the ML model. In order to avoid underfitting and overfitting of the constructed models, a multi-fold cross-validation scheme was selected to build ML models. The following six ML algorithms were selected for training and parameter tuning: support vector machine (SVM), XGBoost, RF, naïve Bayesian (NB), logistic regression (LR) and gradient boosting machines (GBM). The confusion matrix and receiver operating characteristic (ROC) curve were selected to analyse and compare the classification and diagnosis results of the above six ML models.

Parameter settings for model construction

The cross-validation (k-fold) and grid adjustment (GridsearchCv) functions in the Scikit-Learn toolkit were used to optimise the classification parameters of the ML models. The final parameter settings of each algorithm were as follows: (1) SVM: the radial basis function was selected as the kernel function of the SVM model, the penalty coefficient of the SVM model was set to 8 and the gamma value was set to 0.1; (2) construction of the XGBoost model selected the gbtree parameter, and set the learning rate to 0.1; (3) the n_estimators parameter of the RF model was set to 160, and the parameter of max_depth was set to 5; (4) NB: a polynomial model was selected, and its parameters were set to the default value; (5) LR: the penalty item of the model was set to 1.2, and the classification method of ovr was adopted; (6) GBM: the value of n_estimators was set to 100, the parameter settings were set to the default value; (5) LR: the penalty item of the model was set to 1.2, and the classification method of ovr was adopted; (6) GBM: the value of n_estimators was set to 100, the parameter settings were set to the default value.
parameter of learning rate was set to 0.1 and the parameter of max-depth was set to 5.

Evaluation of the comprehensive performance of the models
The accuracy, precision rate, recall rate and F-1 score were used as indicators to evaluate the diagnostic performance of the ML models. The diagnostic accuracy of each model was compared by confusion matrix, and the comparison between models was carried out by ROC curve area. The area of the Precision-Recall curves (PRC) was used to compare the accuracy of each model in classification. F-measure was used as the harmonic average of precision and recall (formula 1), when a=1 in formula 1, F-measure evolved into F-1 (formula 2). Since precision and recall are a pair of contradictory values, the F-1 score is generally used as an evaluation standard to measure the overall performance of the classifier algorithm.

\[
F - \text{measure} = \frac{(a^2+1) \times P \times R}{a^2 (P + R)} \quad (1)
\]

\[
F - 1 = \frac{2P \times R}{P + R} \quad (2)
\]

Statistical analysis
The ML model was constructed by the Python 3.6. The ‘mean function’ in the Python ‘pandas’ library was selected for the mean statistics of measurement data. The t-test and analysis of variance of the measurement data were completed by the ‘std function’ of the pandas library, and the independent t-test of continuous variables was used for the comparison between groups. The quantitative data were expressed in terms of frequency and percentage, and the stats function in the SciPy library was selected to complete the \( \chi^2 \) test of the comparison between different groups. The correlation between sensitivity and specificity was revealed by ROC curve. Each model was evaluated comprehensively by accuracy, precision, recall rate and F-1 score. The p value was calculated by stats_ttest function of SciPy library, and a two-tailed p<0.05 was considered to indicate a statistically significant difference.

RESULTS
Comparison of the performance of three ML algorithms for clinical feature screening
Based on the initial dataset, three different types of ML algorithms were selected for the screening of clinical features, including RF, SelectKBest and XGBoost. The program of each algorithm was run five times, and the average value of the time spent by each algorithm to complete feature screening was used for performance comparison. The results show that the time spent by the three types of algorithms was 2.09±0.14 s, 0.51±0.07 s and 1.85±0.08 s, respectively (figure 1C). There was a significant difference in the time spent by SelectKBest compared with the other two algorithms (95% CI, p<0.01). Based on the performance of the ML algorithm and the consistency of the selected feature items with clinical practice, the SelectKBest algorithm was finally selected to screen the features according to the classification weight and correlation coefficient. The 31-dimensional feature items were selected through the operation of the SelectKBest algorithm and the experimental dataset was established. A total of 359 pieces of NSTEMI data and 342 pieces of UA data were included in the experimental dataset. After dividing the experimental dataset at a ratio of 8:2, 359 pieces of data were used for training and validation of ML models, and 342 pieces of data were used for for testing (figure 2D).

Baseline characteristics of clinical data in experimental dataset
There was no data loss in the process of clinical feature items selection between the training set and the test set. The training set included 476 patients (95%CI 61.1±3.52 years), of which 394 patients (92.55%) were diagnosed with myocardial
infarction by CAG. A total of 225 patients (95% CI 60.7±3.37 years) were included in the test set, of which 209 patients (92.9%) were diagnosed with myocardial infarction by CAG. The age distribution of patients in the two datasets is shown in figure 2A, and the age distribution divided by gender is shown in figure 2B. There was no significant difference between the two datasets in the family history of coronary artery disease (CHD), history of CHD, diabetes mellitus, dyslipidaemia and smoking history (95% CI p=0.37, 0.43, 0.24, 0.39 and 0.15, respectively). There was no significant difference in the proportion of patients whose time from symptom onset to first medical contact (S2FMC) was less than 3 hours in the two data sets (95% CI, p=0.26). According to the Thrombolysis in Myocardial Infarction score, the patients in the two datasets were stratified and there was no significant difference in the proportion of patients with different risks (95% CI high risk, p=0.46; medium risk, p=0.23; low risk, p=0.18). However, there was a significant difference in the proportion of patients with hypertension in the two datasets (95% CI p=0.02). The distribution of hypertension patients divided by cardiac troponin T (cTnT) levels in the two datasets is shown in figure 2C. There was a significant difference in the proportion of patients with grade I and grade II cardiac function of the two datasets (95% CI grade I, p=0.01; grade II, p=0.02) (table 3). The proportion of regional wall motion abnormality in echocardiographic results and the proportion of S2FMC ≤3 hours were not significantly different between the two datasets. According to the number of occurrences of angina pectoris before the visit, figure 2D shows the distribution of age in the two datasets.

Importance ranking of clinical item features screened by SelectKBest

The SelectKBest algorithm was used to screen the clinical feature items in the initial dataset, and the obtained experimental dataset includes 31 feature items. The results of feature screening showed that clinical symptoms, changes in ECG ST-segment and cTnT were included as clinical feature items in the diagnosis. In the ranking results, the top contributing values were cTnT, lactate dehydrogenase, creatine kinase, ST-segment changes in ECG (95% CI 0.21±0.15, 0.11±0.06, 0.08±0.005 and 0.06±0.007, respectively) and other clinical features used in the diagnosis.

### Table 3 Comparison of clinical baseline data between training set and test set

| Variables                        | Training (n=476) | Testing (n=225) | P value |
|----------------------------------|-----------------|----------------|---------|
| Age, years (mean±SD)             | 61.1±3.52       | 60.7±3.37      | 0.19    |
| Sex, female, n (%)               | 45.31           | 44.73          | 0.26    |
| Current smoker, n (%)            | 51.25           | 52.16          | 0.15    |
| Diabetes mellitus                | 34.71           | 32.32          | 0.24    |
| Hypertension, n (%)              | 50.13           | 37.58          | 0.03*   |
| Dyslipidaemia, n (%)             | 41.62           | 39.28          | 0.39    |
| History of CHD (%)               | 28.48           | 27.73          | 0.43    |
| Family history of CAD, n (%)     | 2.15            | 1.94           | 0.37    |
| S2FMC ≤3 hours, n (%)            | 29.91           | 31.14          | 0.26    |
| TIMI score, n (%)                |                 |                |         |
| High risk (5–7)                  | 25.30           | 26.10          | 0.46    |
| Medium risk (3–4)                | 70.50           | 69.90          | 0.23    |
| Low risk (0–2)                   | 4.20            | 4.00           | 0.18    |
| NYHA, n (%)                      |                 |                |         |
| I                                | 78.80           | 65.30          | 0.01*   |
| II                               | 13.70           | 27.20          | 0.02*   |
| III                              | 6.50            | 6.30           | 0.53    |
| IV                               | 1.00            | 1.20           | 0.42    |
| CAG, n (%)                       |                 |                |         |
| CAG (+), n (%)                   | 92.55           | 92.90          | 0.58    |
| CAG (−), n (%)                   | 7.45            | 7.10           | 0.27    |
| RWMA, n (%)                      |                 |                |         |
| RWMA (+), n (%)                  | 43.53           | 45.17          | 0.29    |
| RWMA (−), n (%)                  | 56.47           | 54.83          | 0.33    |
| LVEF                            | 58.65±5.84      | 59.17±4.53     | 0.38    |
| FS (%)                           | 32.48           | 31.93          | 0.49    |

*Comparison of basic clinical baseline data of training set and test set (p<0.05).

CAG, coronary angiography; CHD, coronary artery disease; FS, left ventricular fractional shortening; LVEF, left ventricular ejection fraction; NYHA, New York Heart Association; RWMA, regional wall motion abnormality; S2FMC, symptom to first medical contact; TIMI, Thrombolysis in Myocardial Infarction.
The contribution of each clinical feature was evaluated by plotting shapely values, and the obtained results were consistent with the results of feature contribution ranking.

Concordance and calibration of ML models

The learning curve of the ML model was used to evaluate the performance of the model. This learning curve was used to demonstrate the relationship between sample size and diagnostic accuracy (figure 4A). With the gradual increase of training samples, the models trained by XGBoost, GBM, NB and RF algorithms had better performance and fit in the validation set. The accuracy of the XGBoost model in the validation set was higher than in the other models. Calibration curves were used to describe the agreement between the output probability of each
Figure 5  Confusion matrix diagram for evaluating the classification accuracy of machine learning models. (A) Confusion matrix diagram of each machine learning model. (B) Metrics included in confusion matrix. FN, false negative; FP, false positive; GBM, gradient boosting machines; LR, logistic regression; NB, naive Bayesian; RF, random forest; SVM, support vector machine; TN, true negative; TP, true positive; XGBoost, extreme gradient boosting.
model making an accurate diagnosis and the true probability (figure 4B). The diagnostic consistency of the XGBoost, LR and GBM is better than the other models.

**Construction of confusion matrix of each ML model**

In order to present the visualisation effect of the performance of each ML algorithm and compare the accuracy of classification diagnosis, this study made use of the true-positive (TP), true-negative (TN), false-positive (FP) and false-negative (FN) results in constructing a confusion matrix. The principle of matrix construction is shown in figure 5B, and the confusion matrix diagram of each ML model is shown in figure 5A. According to the results of the confusion matrix, each model exhibited high TP values, indicating excellent positive classification accuracy. However, there are differences in the TN value of each model, and the negative classification accuracy of SVM, XGBoost and LR is better than the other models.

**Construction of ROC and PRC curve of each ML model**

Construction of the ROC curve takes the FP results of each ML model as the abscissa and the TP results as the ordinate. In order to verify the superiority of the ML model, we constructed two types of ROC curves. The first type incorporated the features of cTnT, the number of cases of angina pectoris before the visit and the ST-segment changes of the ECG, while the other type used the features screened in this study. Both options had built ML.

---

**Figure 6** Comprehensive evaluation of the performance of machine learning models. (A) The ROC curves of machine learning models constructed by general clinical features. (B) The ROC curves of the machine learning models constructed using the clinical features screened in this study. (C) The AUC values of the machine learning models were improved by incorporating multiple clinical features (type 1: include general clinical features only, type 2: include the clinical features screened in this study). (D) Schematic diagram of the PRC curves of machine learning models. (E) Display the distribution of the evaluation indicators of each machine learning model through a radar chart. (F) The histogram shows the model coefficient of determination for each machine learning model. GBM, gradient boosting machines; LR, logistic regression; NB, naïve Bayesian; NSTEMI, non-ST-elevation myocardial infarction; RF, random forest; ROC, receiver operating characteristic; SVM, support vector machine; UA, unstable angina pectoris; XGBoost, extreme gradient boosting.
models and plotted the ROC curve (figure 6A, B). Compared with the area under the ROC curve (AUC) of the first type, the AUC values of XGBoost, SVM, RF, GBM and LR model in the second type of feature selection have been improved (95% CI, p=0.003, 0.04, 0.036, 0.002 and 0.041, respectively) (figure 6C). The classification performance of each model with multiple clinical features was better than that of random classifiers. The AUC values of the LR, RF, XGB, GBM and NB were all higher than those of the SVM, and had greater diagnostic accuracy. PRC curves were used to demonstrate the classification performance of the models, in which the AUC values of XGBoost, SVM, LR and RF were higher than the values of other models and had good classification performance (figure 6D). The AUC values of XGBoost and LR models are more balanced in the two types of curves. The comparison of their AUC values in the ROC curve was not significantly different (p=0.31), but there was a significant difference in the PRC curve (p=0.002).

Comprehensive evaluation of each ML model

The performance of the ML models is comprehensively evaluated using accuracy, precision, recall, F-1 score and determination coefficient ($R^2$) (formula 3–5). In the comparison of the comprehensive performance of the models, the XGBoost model outperformed the LR model in terms of accuracy, precision, recall and F-1 score in NSTEMI and UA diagnostic classification. Moreover, in the diagnostic classification of UA, the performance of these indexes of XGBoost is better than that of GBM model (table 4). The radar chart is used to show the balance of the evaluation index distribution of each model, and the XGBoost model outperforms other models (figure 6E). Through the comparison of $R^2$ values, the XGBoost model also showed better fitting performance and interpretability of the results (figure 6F).

\[
\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)
\]

\[
\text{Precision} = \frac{TP}{TP+FP} \quad (4)
\]

\[
\text{Recall} = \frac{TP}{TP+FN} \quad (5)
\]

DISCUSSION

In this study, a database was established by collecting retrospective clinical data of multicentre. Based on this database, six ML models were constructed and used for the auxiliary diagnosis of NSTEMI. Furthermore, a series of evaluation indices were performed to evaluate the accuracy of the models. The comprehensive evaluation of the model showed that the performance of the XGBoost model was better than that of the other models.

In recent years, the application of AI technology in cardiovascular disease has made significant progress. The auxiliary diagnosis provided by the ML model may improve the efficiency of clinical diagnosis, reduce the cost of medical treatment and provide clues for doctors to make decisions during remote diagnosis. This study has a degree of advancement in data processing and feature screening. Clinical data from multiple centres were included in this study, and all contributing centres met all necessary laboratory requirements and guidelines. After completing the preprocessing of the original data, three ML algorithms were selected for feature screening, and the optimal algorithm was finally selected for feature screening through the comparison of the performance of the three algorithms. The SelectKBest algorithm was used to screen clinical feature items to obtain the importance ranking of feature items. The proposed method reduced overfitting, supported column sampling, showed regularisation in the feature screening process, reflected the sequence features more comprehensively and improved the classification. Therefore, the SelectKBest algorithm used in this study also provides clues for the processing and application of clinical data. To avoid bias, this study used a dataset from one medical centre to train the ML models, while a dataset from another medical centre was used to test the models. Not only that, the study also used a fivefold cross-validation scheme to process the dataset to avoid overfitting of the constructed models. In the comparison of clinical baseline data between the segmented training set and test set, there were no significant differences in most cardiovascular disease risk factors. However, in the two datasets, there were significant difference in the proportion of patients with hypertension, as well as patients with heart function grades I and II. The reasons for the above differences might be related to the course of basic disease, S2FMC and medication before visiting doctors. During the process of dataset segmentation and clinical feature screening, the SelectKBest algorithm and fivefold cross-validation selected in this study extracted more effective information from the dataset, and improved the credibility of the ML model built on the dataset.

Based on the same training set, six ML models were constructed in this study, and each model was comprehensively evaluated. The confusion matrix, model accuracy-related indicators and the ROC curve were applied to evaluate the performance of six different models. Based on the same training set data, the AUC values of the XGBoost, SVM, RF,
GBM and LR models incorporating multiple clinical features were improved compared with models that only included features recommended by ACS diagnostic guidelines. Through the analysis and comparison of the results of multiple evaluation indicators, the final results show that the comprehensive performance of the XGB model is better than other models. The early application of XGBoost algorithm focused on the processing of supervised ML data, and it has advantages in the selection of loss of function and optimisation of model algorithms. Khera et al.\textsuperscript{22} used the XGBoost model to predict the risk of death after acute myocardial infarction in a cohort study of 755,402 patients, and the results showed that the XGBoost model performed better than the LR model in risk classification. In a recent study,\textsuperscript{23} XGBoost model was used to identify aberrant blood vessels in CAG images, and showed that the model had high specificity for the exclusion of aberrant vessels. In another study,\textsuperscript{24} it was found that the XGBoost model combined with coronary artery calcification score accurately identified obstructive lesions in the coronary artery. Furthermore, the XGBoost model algorithm constructed by Bertsimas et al.\textsuperscript{25} can accurately predicted seven types of arrhythmic ECG signals. In the field of basic research, the model can also be applied to prediction of subcellular localisation,\textsuperscript{26} lysine glycosylation site\textsuperscript{27} and protein interaction site.\textsuperscript{28} The results of the above studies indicated a good performance of the XGBoost model algorithm in decision-making, explanation and wide applicability. The comprehensive evaluation results of the five models in this study also reflected the advantages of the XGBoost model in the accuracy of NSTEMI diagnosis.

The application of AI has been proved to be beneficial to the diagnosis of cardiovascular diseases\textsuperscript{29,30} and imaging discrimination.\textsuperscript{31} To our knowledge, few studies have systematically compared the application of different ML models for NSTEMI-assisted diagnosis. In summary, the ML models have the potential to be used as an auxiliary diagnostic tool to improve the diagnostic accuracy of NSTEMI and optimise management of patients with NSTEMI. This study contributes new ideas to the risk assessment and clinical decision-making of patients with NSTEMI.

Study limitations
This study had several limitations. This multicentre retrospective study was conducted using clinical data of 2878 patients. Although the quality control scheme was adopted in the data collection, there may still have been subjective selection bias. The clinical data were marked manually by members of the research group, and the errors in the discrimination process may have led to overfitting of the models. The time effect in the course of the disease and the correlation between the characteristic variables were not taken into account. The clinical utility of our model may need to be verified in a prospective cohort to fully assess the accuracy of the diagnosis. Therefore, the repeatability, stability and applicability of the model should be improved in clinical practice.

CONCLUSION
The results of this study support the application of ML models as one of the auxiliary tools to improve the accuracy of NSTEMI diagnosis. Compared with the other four ML models, the comprehensive performance of the XGBoost model had more advantages.

Acknowledgements
The authors gratefully acknowledge the financial supports by the National Natural Science Foundation of China (Grant No. 8660085). Furthermore, we are very grateful to the First Affiliated Hospital of Shihezi University School of Medicine for supporting this study in the work of data collection, and also to the team of the College of Information and Science and Technology of Shihezi University for their help.

Contributors
XM participated in the conception and design of the study. AA collated the data, designed and developed the database. LQ developed the database and contributed to drafting the manuscript. QQ revised it critically for important intellectual content. CXZ completed the data preprocessing. WQH completed the construction of machine learning models. All authors have read and approved the final submitted manuscript. XM is the guarantor for the article who accepts full responsibility for the work and the conduct of the study, had access to the data and oversaw the decision to publish.

Funding
This study was supported by the National Natural Science Foundation of China (Grant No. 8660085).

Competing interests
None declared.

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

Ethics approval
The present study complied with the term of the Declaration of Helsinki and was approved by the Institutional Review Board of Xinjiang Medical

Main messages
- This study used the clinical data of 2878 patients from multiple centres to construct six types of machine learning models to assist in the diagnosis of non-ST-segment elevation myocardial infarction.
- The machine learning model can be used as one of the auxiliary tools to improve the accuracy of non-ST-segment elevation myocardial infarction diagnosis.
- The results of the comprehensive evaluation of the five models revealed that the extreme gradient boosting model has advantages in auxiliary diagnosis.

Current research questions
- Multicentre and prospective studies to confirm the accuracy of the model’s diagnosis may be a new research question.
- The study was conducted at two medical centres in Xinjiang; therefore, it may be necessary to include clinical data from other regions to verify the results of this study.
- Future prospective studies may need to explore the correlation between the time effect of onset and characteristic variables.
- Transform the machine learning models constructed in this study into auxiliary diagnostic tools and apply them in clinical practice.

What is already known on the subject
- The high morbidity and mortality rate of non-ST-segment elevation myocardial infarction makes it particularly important to implement accurate diagnosis in the early stage of the disease.
- According to the current literature, many research results have revealed that machine learning models have high performance in disease risk prediction, image analysis and disease diagnosis.
- However, there are few studies have shown the possibility of using machine learning models as auxiliary tools for the diagnosis of non-ST-segment elevation myocardial infarction.
REFERENCES
1 Collet J-P, Thiele H, Barbato E, et al. 2020 ESC guidelines for the management of acute coronary syndromes in patients presenting without persistent ST-segment elevation. Eur Heart J 2021;42:1289–367.
2 Puymirat E, Simon T, Cayla G, et al. 2022; Postgrad Med J Qin L, et al. doi:10.1136/postgradmedj-2021-141329
3 Jobs A, Collet J-P, Thiele H. Optimal timing of invasive angiography in non-ST-segment elevation acute coronary syndromes: do we need more data? Eur Heart J 2021;42:353–4.
4 Collet J-P, Thiele H. Lesson learnt from the new 2020 ESC guidelines on non-ST-segment elevation acute coronary syndrome: do we need more data? Eur Heart J 2021;42:2607–8.
5 The Writing Committee of the Report on Cardiovascular Health and Diseases in China. Report on cardiovascular health and diseases in China 2019. Chin Circ J 2020;35:833–54.
6 Hoedemaker NPG, Damman P, Woudstra P, et al. Early Invasive Versus Selective Strategy for Non-ST-Segment Elevation Acute Coronary Syndrome: The ICTUS Trial. J Am Coll Cardiol 2017;69:1883–93.
7 Jneid H, Addison D, Bhatt DL, et al. 2017 AHA/ACC clinical performance and quality measures for adults with ST-elevation and non-ST-elevation myocardial infarction: a report of the American College of Cardiology/American heart association task force on performance measures. J Am Coll Cardiol 2017;70:2048–90.
8 Cheng XS, Junbo G. Epidemic of cardiovascular disease in China. Circulation 2018;138:342–4.
9 Fuecker C, Gugala M, Adamson PD, et al. Incidence and outcomes of unstable angina compared with non-ST-elevation myocardial infarction. Heart 2019;105:1423–31.
10 Eggers KM, Jemberg T, Lindahl B. Cardiac Troponin Elevation in Patients Without a Specific Diagnosis. J Am Coll Cardiol 2019;73:1–9.
11 Rashid M, Curzen N, Kinnaird T, et al. Baseline risk, timing of invasive strategy and guideline compliance in NSTEMI: nationwide analysis from MINAP. Int J Cardiol 2020;301:7–13.
12 Berg DD, Wiviott SD, Braunwald E, et al. Modes and timing of death in 66,252 patients with non-ST-segment elevation acute coronary syndromes enrolled in 14 TIMI trials. Eur Heart J 2018;39:3810–20.
13 Than MP, Pickering JW, Sandoval Y, et al. Machine learning to predict the likelihood of acute myocardial infarction. Circulation 2019;140:899–909.
14 Li Y, Yu M, Dai X, et al. Detection of hemodynamically significant coronary stenosis: CT myocardial perfusion versus machine learning CT fractional flow reserve. Radiology 2019;293:305–14.
15 Hou Z-H, Lu B, Li Z-N, et al. Machine learning for pretest probability of obstructive coronary stenosis in symptomatic patients. JACC Cardiovasc Imaging 2019;12:2584–6.
16 Michael W, Daniele M, Andrew D. Deep learning in biomedicine. Nat Biotechnol 2016;36:830–7.
17 Turakhia MP, Desai M, Hedlin H, et al. Rationale and design of a large-scale, app-based study to identify cardiac arrhythmias using a smartphone: the apple heart study. Am J Cardiol 2019;207:66–75.
18 Li W, Yin Y, Quan X, et al. Gene expression value prediction based on XGBoost algorithm. Front Genet 2019;10:1077.
19 Bi Y, Xiang D, Ge Z, et al. An Interpretable Prediction Model for Identifying N7-Methylguanosine Sites Based on XGBoost and SHAP. Mol Ther Nucleic Acids 2020;22:362–72.
20 Mathur P, Sinishta S, Xu X, et al. Artificial intelligence, machine learning, and cardiovascular disease. Clin Med Insights Cardiovasc Med 2020;14:11795682092740–9.
21 Mamuni D, Babu SA. High accurate and a variant of k-fold cross validation technique for predicting the decision tree classifier accuracy. Int J Innov Tech Explor Eng 2021;10:105–10.
22 Khera R, Haimovich J, Hurley NC, et al. Use of machine learning models to predict death after acute myocardial infarction. JAMA Cardiol 2021;6:633–41.
23 Al’Aref SJ, Singh G, Choi JW, et al. A boosted ensemble algorithm for determination of plaque stability in high-risk patients on coronary cta. JACC Cardiovasc Imaging 2020;13:2162–73.
24 Al’Aref SJ, Malakal G, Singh G, et al. Machine learning of clinical variables and coronary artery calcium scoring for the prediction of obstructive coronary artery disease on coronary computed tomography angiography: analysis from the confirm registry. Eur Heart J 2020;41:359–67.
25 Bertsimas D, Mingardi L, Stellato B. Machine learning to predict the likelihood of death after acute myocardial infarction. IEEE J Biomed Health Inform 2021;25:3627–37.
26 Pang L, Wang J, Zhao L, et al. A novel protein subcellular localization method with CNN-XGBoost model for Alzheimer’s disease. Front Genet 2018;9:751–60.
27 Yu L, Shi S, Zhang F, et al. PredGly: predicting lysine glycation sites for Homo sapiens based on XGBoost feature optimization. Bioinformatics 2019;35:2749–56.
28 Deng A, Zhang H, Wang W, et al. Developing computational model to predict protein-protein interaction sites based on the XGBoost algorithm. Int J Mol Sci 2020;21:2274–83.
29 Johnson KW, Torres Soto J, Glicksberg BS, et al. Artificial intelligence in cardiology. J Am Coll Cardiol 2018;71:2668–79.
30 Fernández BI. Artificial intelligence to improve the diagnosis of cardiovascular diseases. Nat Rev Cardiol 2019;16:57–40.
31 Dey D, Slomka PJ, Leeson P, et al. Artificial intelligence in cardiovascular imaging: JACC-state-of-the-art review. J Am Coll Cardiol 2019;73:1317–35.