

Combined Metabolomics and Machine Learning Algorithms to Explore Metabolic Biomarkers for Diagnosis of Acute Myocardial Ischemia

Jie Cao

Shanxi Medical University

Jian Li

Shanxi Medical University

Zhen Gu

Shanxi Medical University

Jia-jia Niu

Shanxi Medical University

Guo-shuai An

Shanxi Medical University

Qiu-xiang Du

Shanxi Medical University

Ke-ming Yun

Shanxi Medical University

Ying-yuan Wang

Shanxi Medical University

Ping Huang

Shanghai Key Laboratory of Forensic Medicine (Academy of Forensic Science)

Jun-hong Sun (✉ junhong.sun@sxmu.edu.cn)

Shanxi Medical University <https://orcid.org/0000-0002-6970-2620>

Research

Keywords: Acute myocardial ischemia, Sudden cardiac death, Metabolomics, Machine learning algorithms, Potential biomarkers

Posted Date: February 4th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-183124/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Version of Record: A version of this preprint was published at International Journal of Legal Medicine on March 29th, 2022. See the published version at <https://doi.org/10.1007/s00414-022-02816-y>.

Abstract

Background: Acute myocardial ischemia (AMI) remains the leading cause of death worldwide. In particular, when death occurs within a short time, it is hard to find post-mortem specific structural anomalies of the heart at autopsy with standard methods. Therefore, the post-mortem diagnosis of AMI represents a current challenge for both clinical and forensic pathologists. Metabolomics technology plays an important role in searching for new diagnostic biomarkers. Here, we characterize metabolic profiles of AMI and attempted to interpret the role of metabolic changes in sudden cardiac death (SCD).

Methods: The untargeted metabolomics was applied to analyze serum metabolic signatures from AMI experimental group (ligation of left coronary artery at 5mm below the left atrial appendage in rats), along with the control and sham groups ($n = 10$ per group). The analytical strategy based on ultra performance liquid chromatography combined with high-resolution mass spectrometry. The resulting data was preprocessed to discriminant metabolites, and a set of machine learning algorithms were used to construct predictable models. Seventeen blood samples from autopsy cases were applied to validate the classification model's value in human samples.

Results: A total of 28 endogenous metabolites in serum were significantly altered in AMI group relative to control and sham groups. Gradient tree boosting, support vector machines, random forests, logistic regression, and multilayer perceptron models were used to further screen the more valuable metabolites from 28 metabolites to optimize the biomarker panel. The results showed that classification accuracy and performance of multilayer perceptron (MLP) models were better than other algorithms when the metabolites consisting of L-threonic acid, N-acetyl-L-cysteine, CMPF, glycocholic acid, L-tyrosine, cholic acid, and glyoursodeoxycholic acid. In autopsy cases, the MLP model constructed based on rat dataset achieved an accuracy of 88.23, and ROC of 0.89 for predicting AMI-SCD.

Conclusions: A panel of 7 molecular biomarkers was identified by assessment the accuracy and efficacy of different metabolite combinations in inferring AMI using machine learning algorithms. The constructed MLP model has a high diagnostic performance for both AMI rats and autopsies-based blood samples. Thus, the combination of metabolomics and machine learning algorithms provides a novel strategy for AMI diagnosis.

Background

Acute myocardial ischemia (AMI) is the primary cause of sudden cardiac death (SCD), which remains a leading cause of morbidity and mortality worldwide[1, 2]. Almost 85% of all the sudden deaths are due to cardiac causes, and many of them asymptomatic at risk of sudden death[3]. AMI can lead the patient to death very quickly, before the appearance of the necrosis at the histological level. Therefore, in majority of such cases, it is hard to find post-mortem specific structural anomalies (both macroscopic and microscopic) of the heart at autopsy with ordinary histological methods, resulting in the cause of death uncertain in practice of forensic pathology[4, 5].

Nowadays, there is a lack of knowledge about the changes that occur shortly after starting the ischemia in humans due to SCD. The molecular autopsy should be considered as a part of the comprehensive medicolegal investigation in SCD cases without structural heart alterations in recent years[6, 7]. However, only 40% of the SCD could be uncovered by molecular autopsy, and these explanations may be more about understanding the genetic correlation of SCD[8, 9]. Until now, there is no highly specific and sensitive "gold standard" for the diagnosis of AMI, and the post-mortem diagnosis of AMI represents a challenge for both clinical and forensic pathologists.

Metabolomics is a powerful approach to detect a broad spectrum of small-molecule metabolites[10, 11], which focuses on studying the metabolic pathway of endogenous metabolites under external and internal effects factors overtime in the organisms, organs, or tissue. With the development of metabolomics technology, it is increasingly being used to study and evaluate cardiovascular diseases and the discoveries of metabolic biomarkers and implications of their causal relationship to cardiovascular disease pathogenesis[12–14]. Ultra performance liquid chromatography combined with high-resolution mass spectrometry (UPLC-HRMS) is considered a better approach for discovering new biomarkers, with high throughput, repeatability, and sensitivity to detect the low concentrations of metabolites.

Coupling with the development in instrumentations, state-of-the-art data analysis tools are needed to handle a large amount of generated metabolite data in untargeted metabolomics. Machine learning algorithms represent potent tools for metabolomics analysis, have been increasingly applied to the classification and data mining of complex UPLC-HRMS data[15]. More recently, the deep learning method or deep belief network (MDBN) in metabolomics has shown the advantages of classifying the different diseases, such as breast cancer status and diagnosing breast hyperplasia[16, 17]. And some reports proved that machine learning algorithms such as support vector machine, genetic algorithm, and random forest could be used for the classification of metabolomic data analysis and provide a high accuracy rate in diagnosing disease[16–18].

In present study, we performed UPLC-HRMS to explore the metabolic characteristics of AMI in the context of sudden death. The multivariate data analysis and a variety of machine learning algorithms, such as gradient tree boosting (GTB), support vector machines (SVM), random forests (RF), logistic regression (LR), and multilayer perceptron (MLP), were used to comprehensively extract the important of disturbed metabolites and assess the performance of machine learning methods on classifying AMI metabolomics data.

Materials And Methods

Chemicals

HPLC-grade acetonitrile (ACN), methanol, and formic acid purchased from Sigma-Aldrich (St. Louis, MO, USA). Deionized water purified through a Milli-Q® purification system from Merck (Millipore, Bedford, MA, USA). Other chemicals, reagents, and solvents used were all of the analytical grades.

Rat experimental protocol

All animal experiments were performed in accordance with the applicable Chinese legislation and approved by the Ethics Committee of Shanxi Medical University, PR China. Sprague-Dawley (SD) rats, weighing 180~220 g, 10~12 weeks old, were supplied by Animal Center of Shanxi Medical University. The rats were housed in cages with rat chow and water under a 12-h light-dark cycle at room temperature (22 to 24°C) and were fasted overnight before the experiment.

All animals were randomly divided into 3 groups ($n = 10$): control group, sham group, and AMI group. The rat model of AMI was established according to conventional coronary ligation[19]. Briefly, rats were anesthetized with intraperitoneal administration of 3% pentobarbital sodium (30 mg/kg), and the lead II electrocardiogram (ECG) was monitored using a BL-420 biological functional experimental system (Chengdu Technology & Market Co. Ltd, China). Then rats were endotracheal intubated and ventilated with a small animal ventilator (HX-100E, Chengdu Technology & Market Co., Ltd, China). A left thoracotomy was performed, and the heart was exposed. The left coronary artery was ligated approximately 5 mm from the lower margin of the left auricle. The left ventricle apex for myocardial blanched and ST segment of ECG was elevated, which indicated that ligation of coronary artery occlusion MI was successful. After 1h of myocardial ischemia, rats were over-anesthetized with pentobarbital sodium to death. Sham-operated rats underwent a similar process without ligation of the left coronary artery. The control group received no treatment.

Blood samples were withdrawn from the rat's abdominal aorta and centrifuged at 12000 rpm for 15 min at 4°C. The supernatant serum samples were aliquoted and immediately stored at -80°C until analysis.

Human blood samples collection by forensic autopsy

This study was approved by the Ethics Committee of Shanxi Medical University, PR China, and all samples were analyzed anonymously. A total of 17 blood samples from forensic autopsy cases, in which 9 cases were confirmed of cardiac cause of death and 8 of noncardiac sudden death in the autopsy. The causes of death of these forensic cases were determined by professional forensic pathologists through systematic forensic autopsies (including macromorphological, histological, toxicological and biochemical examinations) in combination with the death circumstances and medical history of the victims. The forensic autopsies were performed by the Department of Forensic pathology, Shanxi Medical University. Written informed consent statements were acquired from the family member of the deceased individuals.

Sample preparation

Serum samples were thawed before extraction. Then, 800 μ L of cold acetonitrile was added into 200 μ L of serum to remove protein. After vortex mixing for 1 min and centrifugation (12,000 rpm, 20 min, at 4 °C), 600 μ L of the supernatant was withdrawn and freeze-dried in a freeze concentration centrifugal dryer (NingBo XinZhi. Ltd, China). Finally, the residues were dissolved with 200 μ L acetonitrile/water (4:1)

solution, and filtered by 0.22 µm membrane for UPLC-HRMS analysis. A quality control (QC) sample was prepared by pooling and mixing equal-volume sub-aliquots of all samples to monitor the stability of analytical method and system.

UPLC-HRMS analysis of serum samples

The UPLC-HRMS analysis was performed with the Thermo Scientific Ultimate™ 3000 UHPLC system coupled to a Thermo Scientific Q Exactive™ Orbitrap high-resolution mass spectrometer (Thermo Scientific, San Jose, CA, USA) which could acquire the MS² information in a single sample run. Chromatographic separation was performed on an Acquity HSS T3 column (1.8µm, 2.1mm×100 mm, Waters). The column was kept at 40°C, and the injection volume was 5µL. Mobile phase consisted of 0.1% formic acid in water (v/v, A) and 0.1% formic acid in acetonitrile (v/v, B). The flow rate was 0.3 mL/min, with the elution gradient as follows: 0–5 min, 2% B; 5–13 min, 30%B; 13–15 min, 85% B; 15–17min, 98% B; 17–17.5 min, 2% B; and re-equilibration until 20.5 min.

The critical parameters of mass spectrometry detections were performed as follows: capillary temperature was 350 °C, and spray voltages were 3.5 kV and 3.0 kV for positive ion mode and negative ion mode, respectively. The mass scan range was from 80 to 1200 Da. Scanning mode is Full Scan/dd-MS², and the mass resolution was set to 70 000. The resolution is MS Full Scan 35 000 FWHM, MS/MS 17 500 FWHM, NCE is 12.5, 25, and 37.5 eV.

Data preprocessing

The acquired raw data files (.raw) were imported into Compound Discoverer 3.0 (Thermo Fisher, CA, USA) for initial data processing, including peak integration, nonlinear retention time alignment, filtering, matching, *etc.* Simultaneously, the compounds in the serum were annotated. These metabolic discoveries were achieved using a combination of open online databases (mzCloud, HMDB, *etc.*), local databases, and MS/MS metabolites data greatly improves the accuracy of metabolite identification. The final output data includes compound name, retention time, exact mass-to-charge ratio, peak area, *etc.* All data were imported into Excel to normalize the peak area.

Statistical analysis of UPLC-HRMS data

All normalized metabolomic data matrices were imported into SIMCA-P14.0 software (Umetrics, Malmö, Sweden), and multivariate data analysis was carried out. Principal component analysis (PCA) was used to observe general clusters and outliers. Subsequently, the data were subjected to partial least squares-discriminant analysis (PLS-DA) and orthogonal partial least squares-discriminant analysis (OPLS-DA) where models were built and utilized to identify and reveal differential metabolites accountable for the separation between identified groups. Simultaneously, 200 times response permutation testing and *P*-values (*P*-value<0.05) from CV-ANOVA were performed to evaluate the quality of PLS-DA model and OPLS-DA model.

Furthermore, Mann-Whitney U-test ($P < 0.05$) was used to evaluate the differences of metabolites using SPSS 24.0 software (IBM Corp., Armonk, NY, USA). The potential metabolites were selected according to their corresponding variable importance in the projection (VIP) values of these OPLS-DA models and P value of Mann-Whitney U-test.

Machine learning algorithms and feature selection

To screen more important biomarkers and establish the best mathematical classification model for AMI diagnose, we adapted a representative set of five machine learning algorithms that were applied widely in metabolomics: GTB, SVM, RF, LR, and MLP. Before analysis, Z-score data standardization was used to reduce sample variation:

$$Z = (X - \mu) / \sigma$$

Where X is the peak area of each metabolite, μ is the average peak area from each group, $X - \mu$ is the mean deviation, and σ is the standard deviation.

Python software (Intel Corporation, Santa Clara, CA, USA) was employed to develop mathematical models and tune the parameters based on the five machine learning algorithms. The essential features (metabolites) were selected and ranked based on their contribution to each model. Borda count algorithm was applied to summarize all five ranks in order to obtain the final importance rank of metabolites [20]. Ten-fold cross-validation method and average values of area under the curve (AUC) in multivariate receiver operating characteristic (ROC) curve were used to screen the highest performance classification model, and the metabolites in this model were known as the biomarker candidates. The boxplots of biomarkers were prepared using GraphPad Prism7 (GraphPad Software, La Jolla, CA, USA).

Predictive model construction and performance assessment

The best performing model was used as a predicting model for AMI. We randomly split all samples into 70% training set and 30% testing set to assess the overall performance of the model. The 70/30 split is a common practice of splitting ratio for samples of a moderate size in the machine learning applications. Predictive power was assessed by confusion matrices and ROC curves associated with AUC values. Additionally, the metabolomics data of the autopsy cases were used for the external validation set to evaluate the performances of the predicting model for AMI.

Results

Animal model of AMI

We established the AMI animal models by ligation of left coronary artery at 5mm below the left atrial appendage in rats. During the experiment, we observed that a marked elevation in the ST segment of the electrocardiogram (Fig. 1) and myocardium under the ligature went pale, verifying the success of ligation and the occurrence of myocardial ischemia.

Metabolomics statistical analysis of UPLC-HRMS data

As shown in Fig. 2a, QC samples were clustered together in PCA score plots, indicating satisfactory stability and reproducibility of the analysis platform. Due to no visible separation of control, sham, and AMI groups in PCA, the PLS-DA model was sequentially established to explore the differences in metabolic characteristics among each group. There is a good separation in the PLS-DA score plot (Fig. 2b) between the three groups with cumulative R^2Y and Q^2 were 0.987 and 0.814 in the model. Then, the permutation test (200 times, Fig. 2c) and cross-validated residuals analysis of variance (CV-ANOVA) ($P < 0.05$) was utilized to verify the PLS-DA model, and the results showed that there is an adequate capacity for fitting and predicting of the model. Therefore, the PLS-DA analysis demonstrated that the control, sham, and AMI rats could be discriminated against based on metabolic profile.

OPLS-DA was applied to identify the potential metabolites associated with AMI. As illustrated in Fig. 2d, 2e, and 2f, three OPLS-DA models (con vs. sham, con vs. AMI, and sham vs. AMI) were established, and the score plots revealed the separation was effective between groups. Moreover, the results suggest that the VIP scores of metabolites obtained from the OPLS-DA models could be used to confirm the metabolites contributing to separate groups.

Therefore, the differential metabolites were selected according to their corresponding VIP values of the OPLS-DA models and P -value of Mann-Whitney U-test in present study. After eliminating the effects of surgery, 478 differential features with $VIP > 1$ and P -value < 0.05 were screened in the serum of AMI rats (blank circle in Fig. 3a), which attributed to 28 endogenous metabolites. The heatmap was constructed to visualize the difference between the three groups. According to the Figure.3b, there are 22 metabolites up-regulated and 6 metabolites down-regulated compared with control group in serum from AMI rats. The detailed information was presented in Table.1.

Table 1

List of the potential metabolites associated with AMI, which were significantly different (VIP > 1 and *P*-value < 0.05 in serum between AMI and control groups

| Metabolites | Molecular formulae | Molecular weight | Rt(min) | AMI vs Control | |
|--------------------------------|--------------------|------------------|---------|------------------------|----------------------|
| | | | | VIP value ^a | P value ^b |
| L-Tyrosine | C9 H11NO3 | 181.07 | 1.8 | 1.38 | ↓* |
| Stearamide | C18H37NO | 283.29 | 17.36 | 1.12 | ↑* |
| Cholic acid | C24H40O5 | 408.29 | 9.89 | 1.68 | ↓** |
| Glycocholic acid | C26H43NO6 | 465.31 | 9.5 | 1.38 | ↓* |
| L-Ascorbic acid 2-sulfate | C6H8O9S | 255.99 | 1.12 | 1.42 | ↑* |
| 2-Hydroxy-2-methylbutyric acid | C5H10O3 | 118.06 | 4.10 | 1.55 | ↑* |
| Kynurenic acid | C10H7NO3 | 189.04 | 4.59 | 1.67 | ↑** |
| Deoxycholic acid | C24H40O4 | 392.29 | 11.39 | 1.71 | ↓** |
| Glycoursodeoxycholic acid | C26H43NO5 | 449.32 | 10.01 | 1.43 | ↓** |
| N-Acetyl-DL-tryptophan | C13H14N2O3 | 246.1 | 6.65 | 1.31 | ↓* |
| Corticosterone | C21H30O4 | 346.21 | 8.78 | 2.04 | ↑*** |
| Xanthine | C5H4N4O2 | 152.06 | 2.72 | 1.35 | ↑* |
| N3,N4-Dimethyl-L-arginine | C8H18N4O2 | 202.14 | 0.94 | 1.74 | ↑*** |
| N-Acetyl-L-methionine | C7H13NO3S | 191.06 | 4.90 | 2.01 | ↑*** |
| L-Proline | C5H9NO2 | 115.06 | 4.00 | 1.92 | ↑*** |
| 2-Hydroxycaproic acid | C6H12O3 | 132.08 | 5.59 | 1.48 | ↑* |
| DL-4-Hydroxyphenyllactic acid | C9H10O4 | 182.06 | 4.33 | 1.73 | ↑** |
| D-(+)-Galactose | C6H12O6 | 180.05 | 4.53 | 1.39 | ↑* |
| N-Acetyl-L-cysteine | C5H9NO3S | 163.03 | 5.47 | 1.99 | ↑*** |

^a Variable importance in the projection (VIP) value was obtained from OPLS-DA with a threshold of 1.0.

^b *P*-values were derived from Mann-Whitney U-test: **P* < 0.05, ***P* < 0.01, ****P* < 0.001.

Marked with ↑ indicated that the level of metabolites from AMI increased compared with the control group, while marked with ↓ indicated that the level of metabolites decreased.

| Metabolites | Molecular formulae | Molecular weight | Rt(min) | AMI vs Control | |
|--|--------------------|------------------|---------|------------------------|----------------------|
| | | | | VIP value ^a | P value ^b |
| L-Kynurenine | C10H12N2O3 | 208.09 | 3.24 | 1.46 | ↑* |
| L-Threonic acid | C4H8O5 | 136.04 | 0.92 | 1.55 | ↑** |
| 4-Oxoproline | C5H7NO3 | 129.04 | 1.69 | 1.21 | ↑* |
| Uric acid | C5H4N4O3 | 168.03 | 1.49 | 1.14 | ↑* |
| 16-Hydroxyhexadecanoic acid | C16H32O3 | 272.24 | 15.35 | 1.08 | ↑* |
| CMPF | C12H16O5 | 240.1 | 8.19 | 1.53 | ↑* |
| 2-Hydroxyphenylacetic acid | C8H8O3 | 152.05 | 5.86 | 1.79 | ↑** |
| Prostaglandin E1 | C20H34O5 | 354.24 | 9.73 | 1.05 | ↑* |
| 2-Hydroxyhippuric acid | C9H9NO4 | 195.05 | 3.93 | 1.4 | ↑* |
| ^a Variable importance in the projection (VIP) value was obtained from OPLS-DA with a threshold of 1.0. | | | | | |
| ^b <i>P</i> -values were derived from Mann-Whitney U-test: * <i>P</i> < 0.05, ** <i>P</i> < 0.01, *** <i>P</i> < 0.001. | | | | | |
| Marked with ↑ indicated that the level of metabolites from AMI increased compared with the control group, while marked with ↓ indicated that the level of metabolites decreased. | | | | | |

Biomarker candidates and machine learning algorithms optimization for diagnosis of AMI

To extract specific metabolites or groups of metabolites that may represent potential biomarkers directly related to AMI, we adopted a set of machine learning algorithms such as GTB, SVM, RF, LR, and MLP to assess the contribution of metabolites.

The data of 28 metabolites in rat serum were brought into models of GTB, SVM, RF, LR, and MLP algorithms. A feature (metabolite) is considered vital if it contributes to the model performance. The metabolites were ranked according their functional contributions to each model's outputs, respectively (Fig. 4a-e). Every metabolite assigned numerical values in the corresponding model in ascending order, and the smaller score indicated that the contribution is extensive than others. Lastly, Borda count algorithm was applied to summarize all ranks derived from models, and the final importance rank of metabolites was shown in the right column of Fig. 4f.

In order to build the best prediction model by machine learning technique, the new datasets of the different groups of metabolites were derived by removing one by one according to the metabolites' ranks.

According to the ten-fold cross-validation of five machine-learning methods (GTB, SVM, RF, LR, and MLP), the average AUCs of ROC curve analysis were 0.72 for SVM, 0.82 for GTB, 0.76 for RF, 0.82 for LR, 0.98 for MLP, and the accuracy of MLP models achieved 96.67% for the diagnosis of AMI when the groups of metabolites consisting of L-threonic acid, N-acetyl-L-cysteine, CMPF, glycocholic acid, L-tyrosine, cholic acid, and glyoursodeoxycholic acid (Fig. 4g,h).

The above results indicated that the performance of MLP model consisting of these metabolites was better than other models, and the boxplots of normalized intensities for the 7 potential biomarkers were shown in Fig. 5. The levels of 3 of bile acids (cholic acid, glycocholic acid, and glyoursodeoxycholic acid) were both down-regulated in AMI rat serum. The levels of 2 amino acid (threonic acid, N-acetyl-L-cysteine) were up-regulated and L-tyrosine were down-regulated. CMPF was up-regulated in AMI serum.

Validation of the classification model in AMI diagnosis

To better estimate the generalization error, training the parameter of the model, and avoid overfitting, we randomly split all samples into 70% training set and 30% testing set to assess the overall performance of the model. The results of 30% testing set showed that the accuracy of the model was 83.33% and AUC value of ROC curve was 0.88 (Fig. 6a). Confusion matrix of AMI groups and control groups was shown in Fig. 6b, there is only one AMI sample was incorrectly discriminated as the control group.

In present study, a total of 17 human serum samples were collected, including 9 cardiac cause of death and 8 noncardiac sudden deaths confirmed in autopsy (Table 2). The heart blood samples were collected and analyzed by UPLC-HRMS according to the same metabolomics protocol of rat samples. Then, the normalized peak area of the 7 metabolites (L-threonic acid, N-acetyl-L-cysteine, CMPF, glycocholic acid, L-tyrosine, cholic acid, and glyoursodeoxycholic acid) was imported into the constructed MLP model to validate the generalizability of MLP model in human samples. As a result, the accuracy of MLP model which constructed based on rat datasets in autopsies-based blood samples was 88.23%, and AUC value of ROC was 0.89 (Fig. 6c). According to the confusion matrix shown in Fig. 6d, there are only two samples were misjudged into AMI-SCD. The results demonstrated that the MLP model based on the rat metabolomics data achieved a better performance in human samples, and it is a more suitable classifier than other machine learning algorithms for AMI diagnosis.

Table 2
The information of 17 autopsy cases

| | gender | age | actual cause of death | judged by MLP model |
|--|--------|-----|-----------------------|---------------------|
| Case 1 | male | 39 | AMI-SCD | AMI |
| Case 2 | male | 58 | AMI-SCD | AMI |
| Case 3 | female | 50 | AMI-SCD | AMI |
| Case 4 | female | 50 | AMI-SCD | AMI |
| Case 5 | male | 59 | AMI-SCD | AMI |
| Case 6 | male | 53 | AMI-SCD | AMI |
| Case 7 | male | 47 | AMI-SCD | AMI |
| Case 8 | male | 58 | AMI-SCD | AMI |
| Case 9 | male | 63 | AMI-SCD | AMI |
| Case 10 | male | 62 | hemorrhagic shock | no-AMI |
| Case 11 | male | 45 | hemorrhagic shock | no-AMI |
| Case 12 | female | 33 | hemorrhagic shock | no-AMI |
| Case 13 | male | 63 | severe head injury | AMI* |
| Case 14 | male | 66 | severe head injury | AMI* |
| Case 15 | female | 49 | poisoning | no-AMI |
| Case 16 | male | 28 | mechanical asphyxia | no-AMI |
| Case 17 | male | 31 | death from cold | no-AMI |
| *represent misjudged cause of death using MLP model. | | | | |

Discussion

The present study focuses on ischemic heart disease, which represents the most frequent cause of sudden cardiac death in most countries. More remarkably, it is difficult to diagnose in patients who died within 6 hours after the onset of myocardial ischemia because the specific structural anomalies of the heart at autopsy with common method may not have changed. Therefore, the post-mortem diagnosis of AMI represents a current challenge for both clinical and forensic pathologists.

Over the last years, high-sensitivity cardiac troponin T has emerged as biomarker of choice for myocardial damage assessment in the setting of AMI[21, 22]. However, due to the delayed-release kinetics, the clinical value is noticeably limited at the AMI's early stage[23]. Rahimi R *et al.* found that

Troponin T is neither specific nor useful as a cardiac biomarker for post-mortem samples, so it may not be a useful diagnostic tool at autopsy[24, 25]. Therefore, more studies on cardiac biomarkers are needed both in post-mortem practice and clinical diagnose.

The heart has a high metabolic rate to fulfill the demand for adenosine triphosphate (ATP) production to sustain the continual contractile activity. It was reported that in many cardiovascular diseases, the heart undergoes a "metabolic shift", which means the categories or concentrations of metabolites such as fatty acids, glucose, ketone bodies, lactate and amino acids had been changed[26, 27]. Therefore, the alternation of metabolites could help to determine the heart's pathological state and the post-mortem diagnosis of SCD. In present study, we performed a metabolomics protocol to investigate the effect of AMI on the metabolites, and significant differences were detected in metabolic profile between AMI and controls. A total of 28 endogenous metabolites were found disrupted after the AMI event, 22 of them were up-regulated, and 6 of them were down-regulated. The perturbation of the same metabolites occurred similarly in post-mortem blood samples from corpses who died due to AMI.

As described previously, the diagnosis of cause in the first six hours after onset of myocardial ischemia is a great challenge. In this experiment, we investigated the accuracy and efficacy of different metabolite combinations in inferring AMI using five machine learning algorithms. The results showed that the increment of the classification accuracy ranges from 53.33% (SVM, 25 metabolites) to 96.67 (MLP, 7 metabolites), and illustrated that AMI diagnosis can be improved by comparing different small molecule combinations of metabolites and employing appropriate machine learning algorithms (Fig. 4). Furthermore, we used the MLP classification model in 17 forensic autopsies-based blood samples to validate the practicability, and the model achieved an accuracy of 88.23% and ROC of 0.89. These results demonstrated that the 7 metabolites might be the biomarker for clinical diagnosis or post-mortem identification. The "metabolic shift" could as the indicator of the molecular autopsy for medicolegal investigation in SCD cases.

To further understand the relationship between the 7 metabolites and acute myocardial ischemia, we performed a literature review focused on "metabolic shift." Previous studies have reported a relationship between serum bile acid and cardiovascular conditions, including atherosclerosis, obesity, and metabolic diseases[28]. Zhang BC *et al.* found that higher serum total bile acid level was an independent predictor of high-risk coronary plaques in asymptomatic individuals[29], which indicated that serum bile acid level might be forecast the AMI and the occurrence of SCD. It is noteworthy that in the seven essential metabolites, there are three bile acids (cholic acid, glycocholic acid, and glyoursodeoxycholic acid) changes in serum of AMI according present study. These three bile acids significantly altered in AMI showed that the occurrence of myocardial ischemia led to abnormal lipid metabolism, especially bile acid biosynthesis and metabolism.

In our study, significant alterations of threonic acid, N-acetyl-L-cysteine and L-tyrosine were observed in serum associated with amino acid metabolism aberrance of myocardial ischemia. Sun L *et al.* found that the level of threonic acid was markedly down-regulated in the isoproterenol (ISO)-induced AMI rats. It was

considered one of the most associated cecal metabolites with AMI[30]. Nevertheless, we found that the serum level of threonic acid is up-regulated, which may be caused by increased intestinal reabsorption into the blood circulation coupled with oxidative stress response. Thus, the results further confirmed that threonic acid is an important biomarker associated with AMI.

In the present study, the combination of metabolomics and machine learning algorithms showed great potential for diagnosing AMI. However, this study has some limitations. A point of consideration is that larger sample sizes from AMI patients are required in future studies to evaluate the applicability of the AMI model further. Another consideration is that the models were constructed based on animal data, and only 17 autopsies-base blood samples were used for external validation. Thus, large scale human samples are needed to verify the applicability of the diagnosis AMI and SCD in future research.

Conclusions

In summary, there are 28 different metabolites identified by the metabolomics analysis from AMI rats and human serums. Multiple machine learning algorithms identified a panel of 7 highly discriminating metabolites among 28 metabolites to select the biomarkers and validate the diagnostic model. The constructed MLP classification model has a high diagnostic performance for both AMI rats and forensic autopsies-base blood samples. Therefore, the combination of metabolomics and machine learning algorithms can extract important metabolites and has great potential for diagnosing AMI and SCD.

Abbreviations

AMI: Acute myocardial ischemia; SCD: Sudden cardiac death; UPLC-HRMS: Ultra performance liquid chromatography combined with high-resolution mass spectrometry; SVM: Support vector machine; GTB: Gradient tree boosting; RF: Random forest; LR: Logistic regression; MLP: Multilayer perceptron; ECG: Electrocardiogram; PCA: Principal component analysis; PLS-DA: Partial least squares-discriminant analysis; OPLS-DA: Orthogonal partial least squares-discriminant analysis; AUC: Area under the curve; ROC: Receiver operating characteristic.

Declarations

Acknowledgement

Not applicable.

Authors' contributions

JC and JN performed the experiments and wrote the manuscript, JL and GA contributed to data interpretation of data, ZG, KY, QD and PH helped with data acquisition and manuscript modification, and YW and JS designed this research and modified the manuscript. All authors have agreed to the published version of the manuscript. All authors read and approved the final manuscript.

Funding

This research was supported by the Fund from Shanghai Key Laboratory of Forensic Medicine (Academy of Forensic Science) (Grant Numbers KF1803) and the National Natural Science Foundation of China (81901924, 81971795).

Availability of data and materials

The datasets used and/or analyzed during the current study are available from the corresponding author on reasonable request.

Ethics approval and consent to participate

The study was approved by the Ethics Committee of Shanxi Medical University, Shanxi, China.

Written informed consent statements were acquired from family of the deceased.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

References

1. Priori SG, Blomstrom-Lundqvist C, Mazzanti A, et al. 2015 ESC Guidelines for the management of patients with ventricular arrhythmias and the prevention of sudden cardiac death: The Task Force for the Management of Patients with Ventricular Arrhythmias and the Prevention of Sudden Cardiac Death of the European Society of Cardiology (ESC). Endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC). *Eur Heart J*. 2015;36(41):2793-867.
2. Hayashi M, Shimizu W, Albert CM. The spectrum of epidemiology underlying sudden cardiac death. *Circ Res*. 2015;116(12):1887-906.
3. Campuzano O, Allegue C, Partemi S, et al. Negative autopsy and sudden cardiac death. *Int J Legal Med*. 2014;128(4):599-606.
4. Visona SD, Benati D, Monti MC, et al. Diagnosis of sudden cardiac death due to early myocardial ischemia: An ultrastructural and immunohistochemical study. *Eur J Histochem*. 2018;62(2):2866.
5. Aljakna A, Fracasso T, Sabatasso S. Molecular tissue changes in early myocardial ischemia: from pathophysiology to the identification of new diagnostic markers. *Int J Legal Med*. 2018;132(2):425-38.
6. Brion M, Sobrino B, Martinez M, et al. Massive parallel sequencing applied to the molecular autopsy in sudden cardiac death in the young. *Forensic Sci Int Genet*. 2015;18:160-70.

7. Wu J, Wu Q, Dai W, et al. Serum lipid feature and potential biomarkers of lethal ventricular tachyarrhythmia (LVTA) induced by myocardial ion channel diseases: a rat model study. *Int J Legal Med.* 2018;132(2):439-48.
8. Santori M, Blanco-Verea A, Gil R, et al. Broad-based molecular autopsy: a potential tool to investigate the involvement of subtle cardiac conditions in sudden unexpected death in infancy and early childhood. *Arch Dis Child.* 2015;100(10):952-56.
9. Michaud K, Lesta MM, Fellmann F, et al. Molecular autopsy of sudden cardiac death: from postmortem to clinical approach. *Rev Med Suisse.* 2008;4(164):1590-3.
10. Johnson CH, Ivanisevic J, Siuzdak G. Metabolomics: beyond biomarkers and towards mechanisms. *Nat Rev Mol Cell Biol.* 2016;17(7):451-9.
11. Cao J, Jin Q Q, Wang G M, et al. Comparison of the serum metabolic signatures based on (1)H NMR between patients and a rat model of deep vein thrombosis. *Sci Rep.* 2018;8(1):7837.
12. McGarrah R W, Crown S B, Zhang G F, et al. Cardiovascular Metabolomics. *Circ Res.* 2018;122(9):1238-58.
13. Au A. Metabolomics and Lipidomics of Ischemic Stroke. *Adv Clin Chem.* 2018;85:31-69.
14. Li Y, Zhang D, He Y, et al. Investigation of novel metabolites potentially involved in the pathogenesis of coronary heart disease using a UHPLC-QTOF/MS-based metabolomics approach. *Sci Rep.* 2017;7(1):15357.
15. Heinemann J. Machine Learning in Untargeted Metabolomics Experiments[M]//New York, NY: Springer New York. 2018:287-99.
16. Jiang M, Liang Y, Pei Z, et al. Diagnosis of Breast Hyperplasia and Evaluation of RuXian-I Based on Metabolomics Deep Belief Networks. *Int J Mol Sci.* 2019;20(11).
17. Alakwaa FM, Chaudhary K, Garmire LX. Deep Learning Accurately Predicts Estrogen Receptor Status in Breast Cancer Metabolomics Data. *Journal of proteome research.* 2018;17(1):337-47.
18. Wang X, Chen H, Chang C, et al. Study the therapeutic mechanism of Amomum compactum in gentamicin-induced acute kidney injury rat based on a back propagation neural network algorithm. *J Chromatogr B Analyt Technol Biomed Life Sci.* 2017;1040:81-8.
19. Gao Y, Gao J, Chen C, et al. Cardioprotective effect of polydatin on ventricular remodeling after myocardial infarction in coronary artery ligation rats. *Planta Med.* 2015;81(7):568-577.
20. Mamoshina P, Volosnikova M, Ozerov IV, et al. Machine Learning on Human Muscle Transcriptomic Data for Biomarker Discovery and Tissue-Specific Drug Target Identification. *Front Genet.* 2018;9:242.
21. Tiller C, Reindl M, Holzknrecht M, et al. Biomarker assessment for early infarct size estimation in ST-elevation myocardial infarction. *Eur J Intern Med.* 2019;64:57-62.
22. Reinstadler SJ, Feistritz HJ, Klug G, et al. High-sensitivity troponin T for prediction of left ventricular function and infarct size one year following ST-elevation myocardial infarction. *Int J Cardiol.* 2016;202:188-93.

23. Laugaudin G, Kuster N, Petiton A, et al. Kinetics of high-sensitivity cardiac troponin T and I differ in patients with ST-segment elevation myocardial infarction treated by primary coronary intervention. *Eur Heart J Acute Cardiovasc Care*. 2016;5(4):354-363.
24. Rahimi R, Dahili ND, Anuar ZK, et al. Post mortem troponin T analysis in sudden death: Is it useful? *Malays J Pathol*. 2018;40(2):143-8.
25. Beausire T, Faouzi M, Palmiere C, et al. High-sensitive cardiac troponin hs-TnT levels in sudden deaths related to atherosclerotic coronary artery disease. *Forensic Sci Int*. 2018;289:238-43.
26. van der Vusse GJ, Glatz JF, Stam HC, et al. Fatty acid homeostasis in the normoxic and ischemic heart. *Physiol Rev*. 1992;72(4):881-940.
27. Heather LC, Wang X, West JA, et al. A practical guide to metabolomic profiling as a discovery tool for human heart disease. *J Mol Cell Cardiol*. 2013;55:2-11.
28. Steiner C, Othman A, Saely CH, et al. Bile acid metabolites in serum: intraindividual variation and associations with coronary heart disease, metabolic syndrome and diabetes mellitus. *PLoS One*. 2011;6(11):e25006.
29. Zhang BC, Chen JH, Xiang CH, et al. Increased serum bile acid level is associated with high-risk coronary artery plaques in an asymptomatic population detected by coronary computed tomography angiography. *J Thorac Dis*. 2019;11(12):5063-70.
30. Sun L, Jia H, Li J, et al. Cecal Gut Microbiota and Metabolites Might Contribute to the Severity of Acute Myocardial Ischemia by Impacting the Intestinal Permeability, Oxidative Stress, and Energy Metabolism. *Front Microbiol*. 2019;10:1745.

Figures

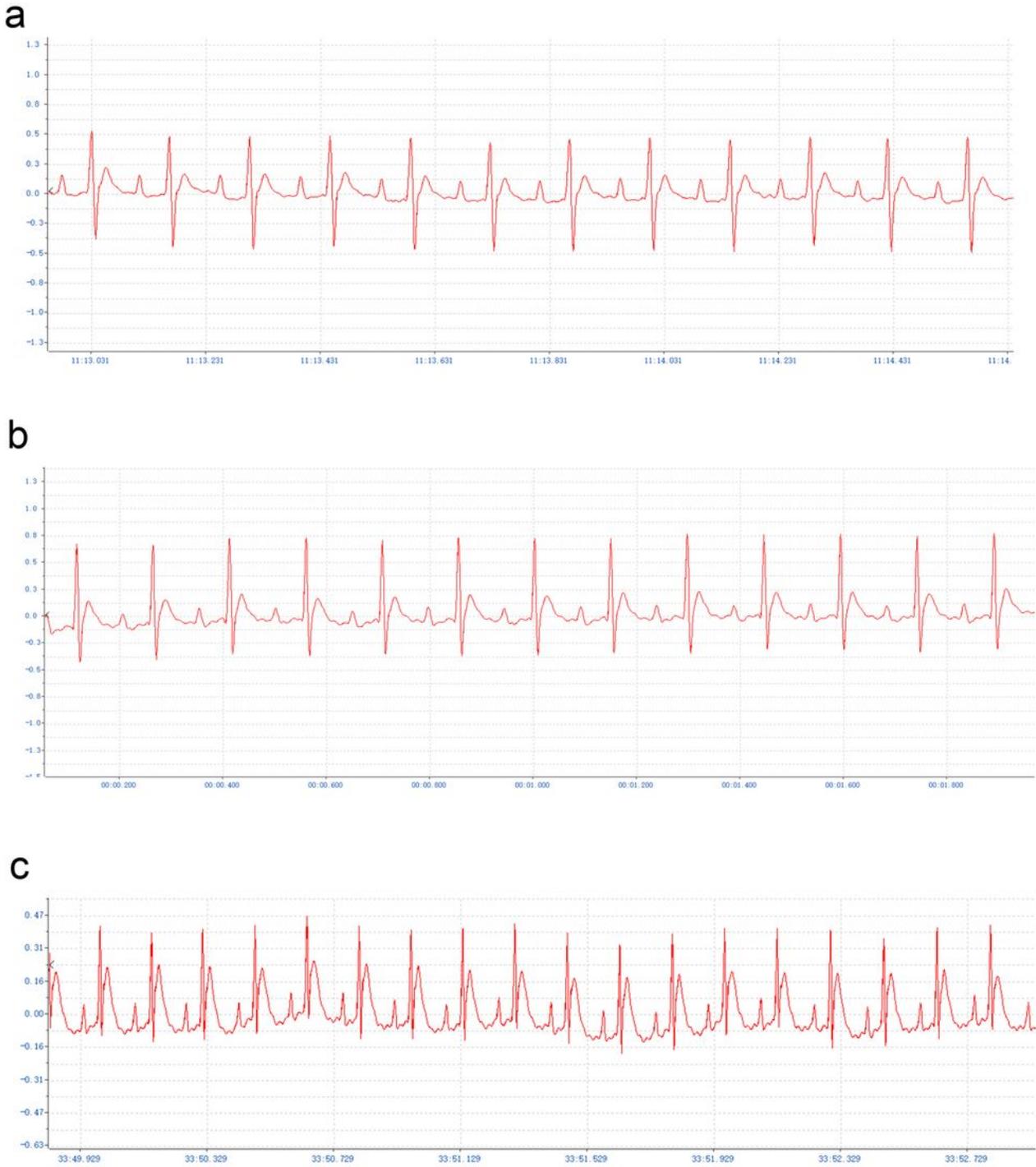


Figure 1

Lead \square ECG electrocardiogram in each rat group. a Normal ECGs of control group; b normal ECGs of sham group; c ECGs of AMI group exerts marked elevation in the ST segment.

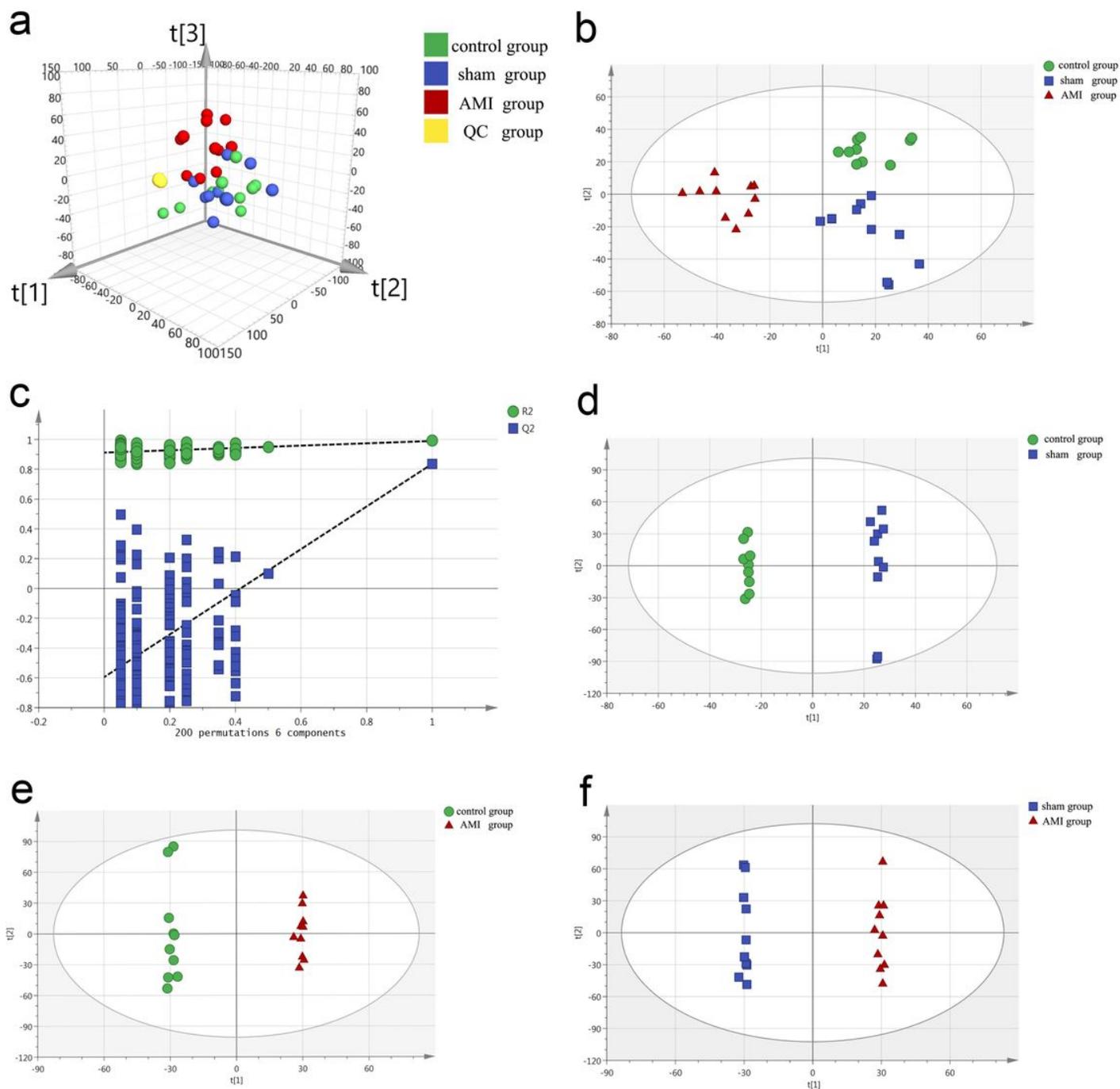


Figure 2

Multivariate analysis. a The PCA score plot of control group (green), sham group (blue) and AMI group (red) revealed a close clustering of quality control (QC) samples (yellow). b PLS-DA score plot suggests that it is possible to discriminate AMI groups from sham and control groups. c The permutation test (200 times) shows that no over-fitting of PLS-DA. d OPLS-DA score plot between control and AMI group. e OPLS-DA score plot between control and sham group. f OPLS-DA score plot between sham and AMI group.

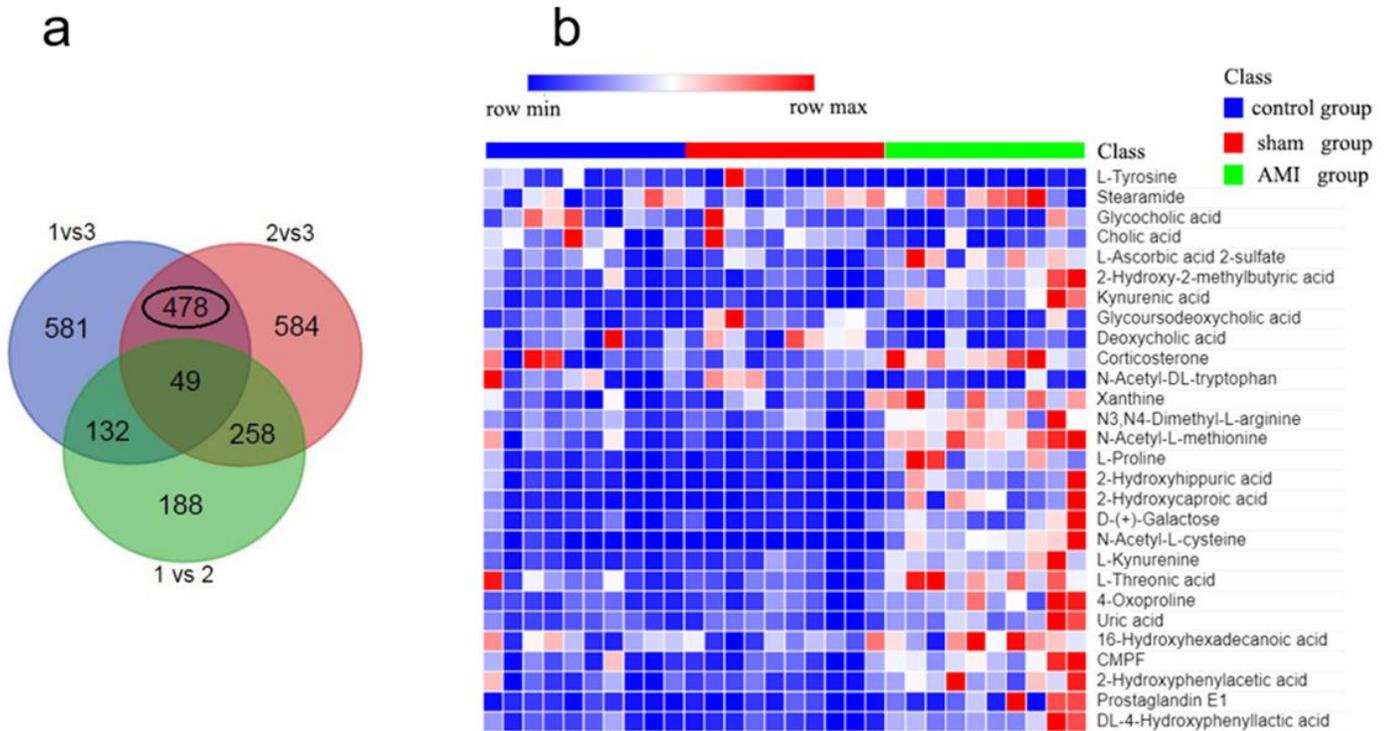


Figure 3

Venn diagram and heatmap plot from the comparisons of AMI, sham, and control groups. a Venn diagram of differential features. The numbers in the overlapping areas (encircled in blank) represent differential features between AMI group and the other two groups, excluding the effect of operation. b Heatmap plot of the 28 potential metabolites. Red color indicates a higher level and green color indicates a lower level.

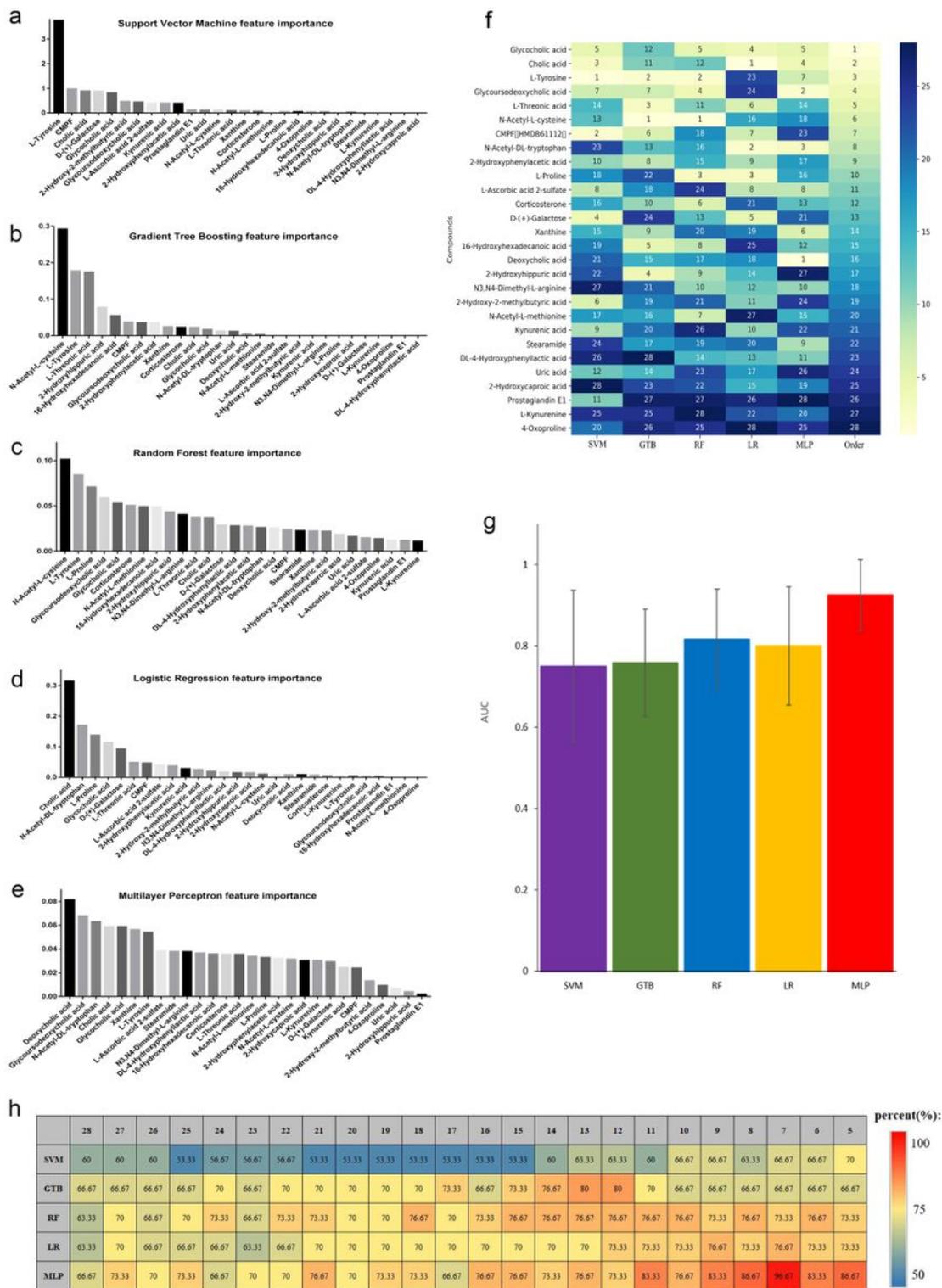


Figure 4

List of the most important metabolites by machine learning algorithms. Histogram of ranking metabolites in support vector machine (SVM) (a), gradient tree boosting (GTB) (b), random forest (RF) (c), logistic regression (LR) (d) and multilayer perceptron (MLP) (e). Heatmap of metabolites' order obtained by the Borda count algorithm (f). Accuracy heatmap of machine learning algorithms Green color indicates a high contribution to the model and blue indicates a low contribution. g The average AUCs of

different machine learning algorithms. h The classification accuracy of the different metabolites' groups and machine learning algorithms for the diagnosis of AMI.

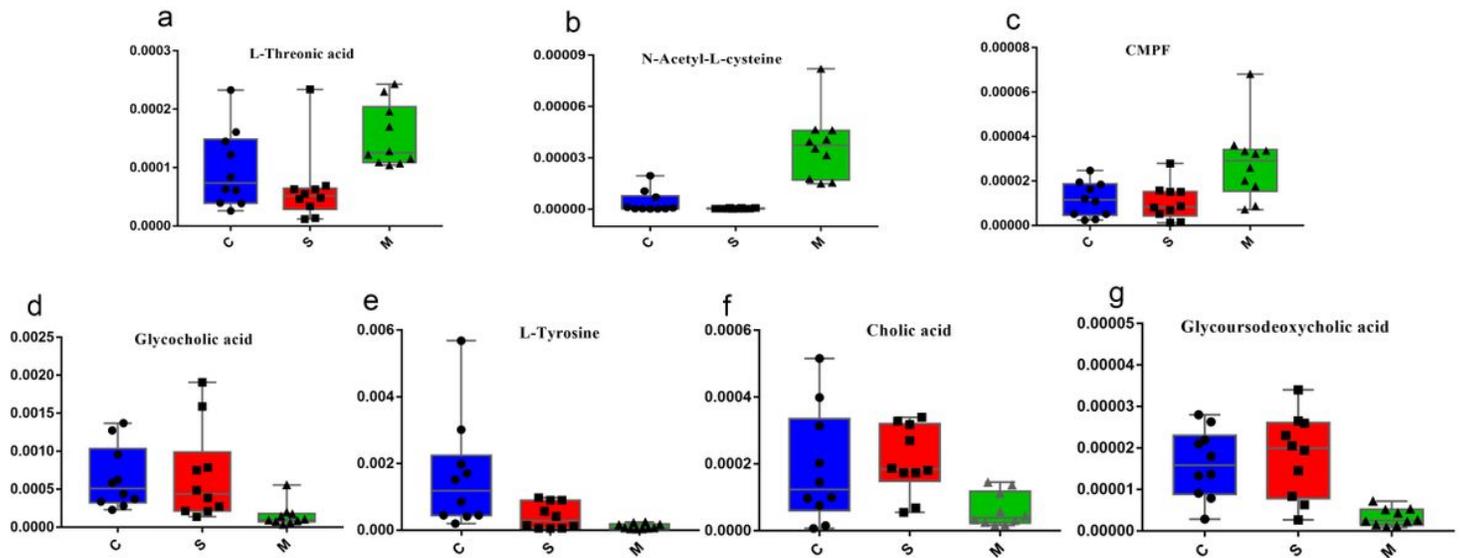
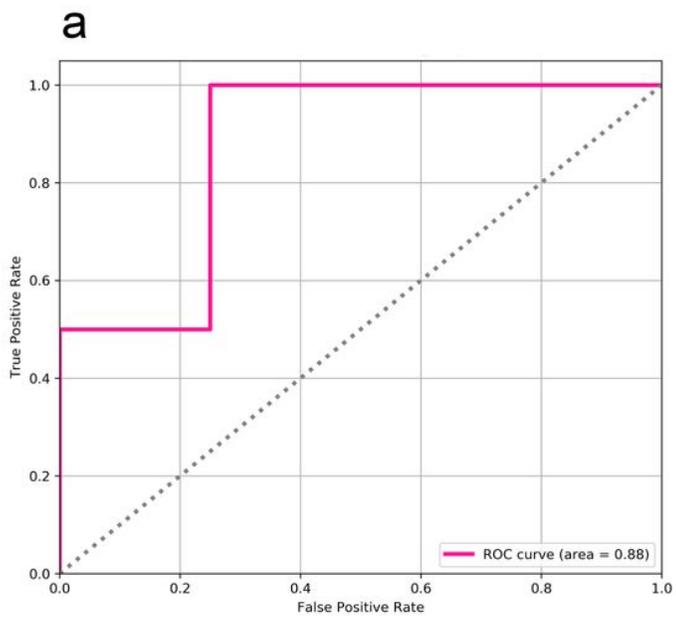


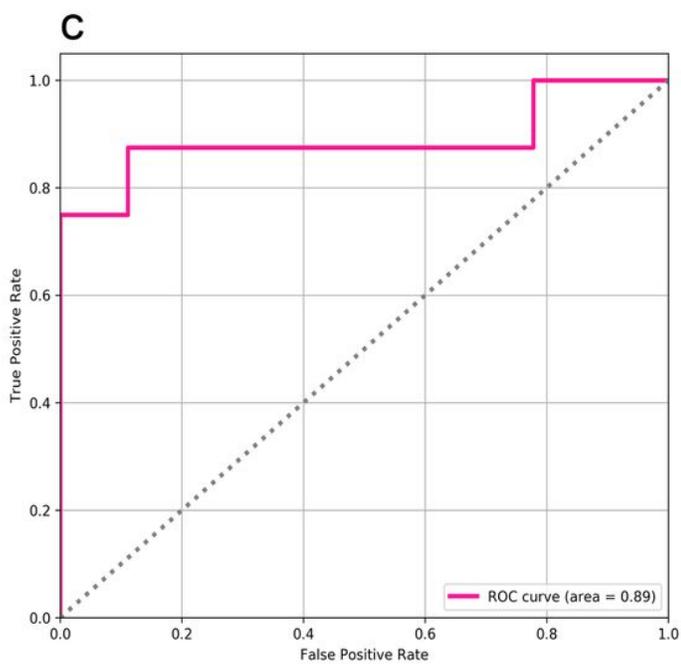
Figure 5

Boxplots of the 7 potential biomarkers in control (C), sham (S) and AMI (M) rat groups, showing a significantly difference level between the AMI and other groups.



b

| Predicted \ Actual | AMI | control |
|--------------------|-----|---------|
| AMI | 9 | 1 |
| control | 0 | 10 |



d

| Predicted \ Actual | AMI | control |
|--------------------|-----|---------|
| AMI | 9 | 0 |
| control | 2 | 6 |

Figure 6

a Receiver operator characteristic (ROC) curve of MLP predicting model in rat samples. b Confusion matrix of AMI groups and control groups in rat. c ROC curve of MLP predicting model in human samples. d Confusion matrix of AMI groups and control groups in human.