Combined metabolomics and machine learning algorithms to explore metabolic biomarkers for diagnosis of acute myocardial ischemia

Jie Cao1,2 · Jian Li2 · Zhen Gu2 · Jia-jia Niu2 · Guo-shuai An2 · Qian-qian Jin2 · Ying-yuan Wang2 · Ping Huang1 · Jun-hong Sun1,2

Received: 8 February 2021 / Accepted: 15 March 2022 / Published online: 29 March 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
Acute myocardial ischemia (AMI) remains the leading cause of death worldwide, and the post-mortem diagnosis of AMI represents a current challenge for both clinical and forensic pathologists. In the present study, the untargeted metabolomics based on ultra-performance liquid chromatography combined with high-resolution mass spectrometry was applied to analyze serum metabolic signatures from AMI in a rat model (n = 10 per group). A total of 28 endogenous metabolites in serum were significantly altered in AMI group relative to control and sham groups. A set of machine learning algorithms, namely gradient tree boosting (GTB), support vector machine (SVM), random forest (RF), logistic regression (LR), and multilayer perceptron (MLP) models, was used to screen the more valuable metabolites from 28 metabolites to optimize the biomarker panel. The results showed that classification accuracy and performance of MLP model 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 glycoursodeoxycholic acid. Finally, 17 blood samples from autopsy cases were applied to validate the classification model’s value in human samples. The MLP model constructed based on rat dataset achieved accuracy of 88.23%, and ROC of 0.89 for predicting AMI type II in autopsy cases of sudden cardiac death. The results demonstrated that MLP model based on 7 molecular biomarkers had a good diagnostic performance for both AMI rats and autopsy-based blood samples. Thus, the combination of metabolomics and machine learning algorithms provides a novel strategy for AMI diagnosis.

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

Introduction
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. Specific markers of early myocardial ischemia, such as fibronectin, C5b-9, troponins, myoglobin, and S100A1, have been effectively used in clinical practice, but lost specificity in human post-mortem samples [6–8]. The molecular autopsy should be considered as a part of the comprehensive medicolegal investigation in SCD cases without structural heart alterations in recent years [9, 10]. 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 [11, 12]. 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 promising tool to improve current, single biomarker-based approaches by identifying metabolic biosignatures that embody global biochemical changes in disease [13, 14]. In terms of AMI, alternations in metabolism are the most immediate molecular changes affected by myocardial ischemia, which can then lead to different deleterious consequences, such as arrhythmia, myocardial infarction, and heart failure [15]. In recent years, the metabolomics technology has been increasingly used to study and evaluate cardiovascular diseases and the discoveries of metabolic biomarkers and implications of their causal relationship to cardiovascular disease pathogenesis [16–18]. Khan A et al. identified that upregulated L-homocysteine sulfinic acid, cysteic acid, and carnitine can help in detecting the subjects who are at risk of developing AMI [19]. Lactate, glutamine, and glutamate, as an adjuvant diagnostic avenue of cardiac troponin for myocardial infarction, have been studied in infarcted myocardia and serum post myocardial ischemia [20]. Compared with myocardial tissues, the biomarkers for diagnosis should enter the serum from the tissues and the metabolites in serum were less affected by the different regions of the heart. Therefore, we infer that molecular biomarkers in serum could be used for the autopsy diagnosis of SCD induced by AMI.

In the present study, we performed ultra-performance liquid chromatography combined with high-resolution mass spectrometry (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 forest (RF), logistic regression (LR), and multilayer perceptron (MLP), were used to comprehensively extract the importance 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 per group): control, sham, and AMI. The rat model of AMI was established according to conventional coronary ligation [21]. 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 ventricular apex of myocardium became pale and ST segment of ECG was elevated, which indicated that occlusion of coronary artery successfully induced myocardial ischemia. After 1 h of myocardial ischemia, the animals were euthanized by lethal dose of sodium pentobarbital. 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 12,000 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 was collected from right ventricle in the forensic autopsy cases, in which 9 cases were confirmed of cardiac cause of death with AMI type II and 8 of noncardiac sudden death in the autopsy (Table 1). The heart blood samples were collected and analyzed by UPLC-HRMS according to the same metabolomics protocol of rat samples. 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 μL of cold acetonitrile was added into 200 μL of serum to remove protein. After vortex mixing for 1 min and centrifugation (12,000 rpm, 20 min, at 4 °C), 600 μL 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 μL 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 (1.8 μm, 2.1 mm × 100 mm, Waters). The column was kept at 40°C and the injection volume was 5 μL. 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; 15–17 min, 98% B; 17–17.5 min, 2% B; and re-equilibration until 20.5 min.

The critical parameters of mass spectrometry detection 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 full width half maximum (FWHM), MS/MS 17,500 FWHM, normalized collisional energy (NCE) values were 12.5, 25, and 37.5 eV.

| Table 1 The information of 17 autopsy cases | Gender | Age | Postmortem interval | Myocardial fibrosis | Actual cause of death | Judged by MLP model |
|--------------------------------------------|--------|-----|---------------------|--------------------|----------------------|---------------------|
| Case 1          Male 39 | 1.8 d  | √   | AMI-SCD             | AMI                | AMI                  |
| Case 2          Male 58 | 6.0 d  | √   | AMI-SCD             | AMI                | AMI                  |
| Case 3          Female 50| 2.4 d  | √   | AMI-SCD             | AMI                | AMI                  |
| Case 4          Female 50| 2.6 d  | √   | AMI-SCD             | AMI                | AMI                  |
| Case 5          Male 59 | 5.0 d  | √   | AMI-SCD             | AMI                | AMI                  |
| Case 6          Male 53 | 6.0 d  | x   | AMI-SCD             | AMI                | AMI                  |
| Case 7          Male 47 | 1.0 d  | √   | AMI-SCD             | AMI                | AMI                  |
| Case 8          Male 58 | 3.0 d  | x   | AMI-SCD             | AMI                | AMI                  |
| Case 9          Male 63 | 4.9 d  | √   | AMI-SCD             | AMI                | AMI                  |
| Case 10         Male 62 | 4.6 d  | x   | Hemorrhagic shock   | no-AMI             | no-AMI               |
| Case 11         Male 45 | 1.0 d  | x   | Hemorrhagic shock   | no-AMI             | no-AMI               |
| Case 12         Female 33| 1.7 d  | x   | Hemorrhagic shock   | no-AMI             | no-AMI               |
| Case 13         Male 63 | 5.3 d  | x   | Severe head injury  | AMI*               | AMI                  |
| Case 14         Male 66 | 3.1 d  | x   | Severe head injury  | AMI*               | AMI                  |
| Case 15         Female 49 | 5.7 d  | x   | Poisoning           | no-AMI             | no-AMI               |
| Case 16         Male 28 | 7.0 d  | x   | Mechanical asphyxia | no-AMI             | no-AMI               |
| Case 17         Male 31 | 4.3 d  | x   | Death from cold     | no-AMI             | no-AMI               |

*Represents misjudged cause of death using MLP model
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, and matching (mass tolerance 5 ppm). Simultaneously, the compounds in the serum were annotated. These metabolic discoveries were achieved by a combination of open online databases (mzCloud, ChemSpider, PubChem and Human Metabolome Database), using exact mass and MS/MS spectra to improve 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 for normalization by the total sum of all compounds’ peak area in that sample, which are also present in all other samples in the experiment.

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. According to the previous study [14], 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 cross-validated residuals analysis of variance (CV-ANOVA) were performed to evaluate the quality of PLS-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 biomarker candidates selection

To screen more important biomarkers and establish the best mathematical classification model for AMI diagnosis, we adapted a representative set of 5 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:

\[ Z = \frac{(X - \mu)}{\sigma} \]

where \( X \) is the peak area of each metabolite, \( \mu \) is the average peak area from each group, \( X-\mu \) is the mean deviation, and \( \sigma \) 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 5 machine learning algorithms. The essential metabolites were selected and ranked based on their contribution to each model. Borda count algorithm was applied to summarize all 5 ranks in order to obtain the final importance rank of metabolites [22]. Tenfold 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 [23]. Additionally, the metabolomics data of the autopsy cases were used for the external validation set to evaluate the performance of the predicting model for AMI.

Results

Animal model of AMI

We established the AMI animal models by ligation of left coronary artery at 5 mm below the left atrial appendage in rats. During the experiment, we observed a marked elevation in the ST segment of the electrocardiogram after ligation (Supplementary Fig. S1) and myocardium under the ligature went pale, verifying the success of ligation and the occurrence of myocardial ischemia.

Cardiac gross and histopathology changes of SCD

For 9 cases of SCD (Table 1), there was obvious thickening of the coronary artery wall and stenosis of the lumen in both autopsy and histopathological examination (Figs. 1a
and 1b). Gray-white scar tissue could be seen in the myocardium of the region dominated by the narrow coronary artery (Fig. 1c), which was replaced by many fibrous tissues under the microscope (Fig. 1d). Pathological manifestations of early myocardial infarction were not found in these cases, such as wave-like changes of cardiac muscle fibers, contraction band necrosis, hemorrhage, and even inflammatory cell infiltration.

Metabolomics statistical analysis of UPLC-HRMS data

As shown in Fig. 2a, QC samples were clustered together in PCA score plot, indicating satisfactory stability and reproducibility of the analysis platform. The PLS-DA score plot (Fig. 2b) illustrated completely separations between the 3 groups with cumulative $R^2_Y$ and $Q^2$ were 0.987 and 0.814 in the model. The results of the permutation test (200 times, Fig. 2c) and CV-ANOVA ($P < 0.05$) showed an adequate capacity for fitting and predicting of the model. Furthermore, the score plots of OPLS-DA models (control vs. sham, control vs. AMI, and sham vs. AMI) revealed that the separation was effective between groups (Figs. 2d, 2e, and 2f). However, the unbiased PCA results exerted a poor separation in 3 groups of metabolomics data (Fig. 2a), especially in control vs sham rats (Fig. 2g), but Figs. 2h and 2i showed good separations in control vs AMI rats and sham vs AMI rats. Based on these results, subsequent comparative analysis was done between ischemic samples and controls.

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 (red circle in Fig. 3a), which attributed to 28 endogenous metabolites. The remaining 450 features did not match any known compounds (Supplementary Table S1). According to the heatmap (Fig. 3b), there are 22 metabolites up-regulated and 6 metabolites down-regulated compared with control group in serum from AMI rats (Table 2).

Biomarker candidates and machine learning algorithms optimization for diagnosis of AMI

To extract more important metabolites, the data of 28 metabolites in rat serum were brought into models of GTB, SVM, RF, LR, and MLP algorithms. A metabolite is considered vital if it contributes to the model performance. The metabolites were ranked according to their functional contributions to each model’s outputs, respectively (Figs. 4a–e). In each model, every metabolite was assigned a numerical score, which indicated its contribution: smaller value represents a more extensive contribution of metabolite to a given model. 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. The new datasets from different groups of metabolites, which were generated by removing one by one according to the metabolites’ ranks, were used to build prediction model. According to the tenfold cross-validation of 5
machine-learning methods, 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 glycoursodeoxycholic acid (Figs. 4g and 4h). The representative chromatograms of 7 different metabolites in the control, sham, and AMI groups were shown in Supplementary Fig. S2.

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.
Table 2  List of the potential metabolites associated with AMI in rat serum

| No | m/z   | Metabolites                   | Molecular formula | Mass error (ppm) | Rt (min) | AMI vs Sham | AMI vs Control |
|----|-------|-------------------------------|-------------------|------------------|---------|-------------|---------------|
| 1  | 180.0669 | L-Tyrosine                  | C9H11NO3          | 3.9              | 1.80    | 1.52        | 1.38          | 0.19          |
| 2  | 284.2951 | Stearamide                  | C18H37NO          | 1.1              | 17.36   | 1.57        | 2.01          | 1.12          | 1.59          |
| 3  | 407.2811 | Cholic acid                  | C24H40O5          | 2.0              | 10.56   | 1.58        | 0.36          | 1.68          | 0.41          |
| 4  | 464.3027 | Glycocholic acid            | C26H43NO6         | 2.2              | 9.50    | 1.63        | 0.46          | 1.38          | 0.38          |
| 5  | 524.9821 | N-Ascorbic acid 2-sulfate   | C6H8O9S           | 2.0              | 1.12    | 1.25        | 1.51          | 1.42          | 1.64          |
| 6  | 117.0588 | 2-Hydroxy-2-methylbutyric acid | C5H10O3         | 0.9              | 4.10    | 1.41        | 2.36          | 1.55          | 3.72          |
| 7  | 190.0501 | Kynurenic acid             | C10H7NO3          | 1.1              | 4.59    | 1.33        | 2.86          | 1.67          | 4.18          |
| 8  | 391.2863 | Deoxycorticosterone       | C24H40O4          | 2.3              | 11.39   | 1.42        | 0.43          | 1.71          | 0.38          |
| 9  | 448.3081 | Glycocholic acid            | C26H43NO5         | 2.7              | 10.68   | 1.68        | 0.68          | 1.43          | 0.21          |
| 10 | 245.0936 | N-Acetyl-DL-tryptophan     | C13H14N2O3        | 1.6              | 6.65    | 1.45        | 0.35          | 1.31          | 0.44          |
| 11 | 347.2218 | Corticosterone             | C21H30O4          | 0.3              | 8.78    | 1.42        | 2.06          | 2.04          | 2.83          |
| 12 | 151.0262 | Xanthine                    | C5H4N4O2          | 0.7              | 2.72    | 1.45        | 2.29          | 1.35          | 1.73          |
| 13 | 203.1505 | N3,N4-Dimethyl-L-arginine  | C8H18N4O2         | 1.0              | 0.94    | 1.76        | 2.30          | 1.74          | 2.65          |
| 14 | 190.0547 | N-Acetyl-L-methionine      | C7H13NO3S         | 2.1              | 4.90    | 1.43        | 2.08          | 2.01          | 5.01          |
| 15 | 116.0707 | L-Proline                   | C5H9NO2           | 0.9              | 4.00    | 1.77        | 4.51          | 1.92          | 7.03          |
| 16 | 131.0715 | 2-Hydroxypropiolic acid    | C6H12O3           | 0.8              | 5.59    | 1.34        | 5.99          | 1.48          | 10.81         |
| 17 | 181.051 | 2-Hydroxyphenylactic acid  | C9H10O4           | 2.2              | 4.33    | 1.32        | 2.21          | 1.73          | 2.83          |
| 18 | 179.0575 | D-(-)-Galactose            | C6H12O6           | 2.2              | 4.53    | 1.33        | 4.95          | 1.39          | 6.49          |
| 19 | 162.0232 | N-Acetyl-L-cysteine        | C5H9NO3S          | 1.2              | 5.46    | 1.87        | 10.64         | 1.99          | 80.32         |
| 20 | 209.0921 | L-Kynurenine               | C10H12N2O3        | 0.0              | 3.24    | 1.12        | 1.90          | 1.46          | 2.20          |
| 21 | 135.0302 | L-Threonic acid            | C4H8O5            | 2.2              | 0.90    | 1.15        | 1.62          | 1.55          | 2.55          |
| 22 | 128.0354 | 4-Oxoproline               | C5H7NO3           | 0.8              | 1.69    | 1.31        | 2.48          | 1.21          | 2.20          |
| 23 | 169.0357 | Uric acid                  | C5H4N4O3          | 0.6              | 1.49    | 1.21        | 2.21          | 1.14          | 2.13          |
| 24 | 271.2284 | 16-Hydroxyhexadecanoic acid | C16H32O3       | 1.8              | 15.35   | 1.02        | 1.43          | 1.08          | 1.57          |
| 25 | 239.0929 | CMPF                        | C12H16O5          | 1.7              | 8.19    | 1.40        | 2.24          | 1.53          | 2.84          |
| 26 | 151.0402 | 2-Hydroxyphenylactic acid  | C8H8O3            | 0.7              | 5.86    | 1.27        | 1.93          | 1.79          | 3.12          |
| 27 | 353.234 | Prostaglandin E1           | C20H34O5          | 2.0              | 9.73    | 1.06        | 3.99          | 1.05          | 4.03          |
| 28 | 194.0462 | 2-Hydroxyhippuric acid     | C9H9NO4           | 1.5              | 3.93    | 1.27        | 3.67          | 1.4           | 4.42          |

\(^{a}\)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
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 AMI in patients who died within 6 h after the onset of myocardial ischemia because of lacking the specific structural anomalies of the heart based on the current methods of histological examination at autopsy. Therefore, the post-mortem diagnosis of AMI represents a current challenge for both clinical and forensic pathologists.

The heart has a high metabolic rate to fulfill the demand for adenosine triphosphate (ATP) production to sustain the continual contractile activity. 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 [24, 25]. In this study, we applied a metabolomics protocol to investigate the effects of AMI on the metabolites and identified early biomarkers of AMI that can distinguish between AMI and control rats. Compared with the other published AMI metabolomics studies mainly engaged in clinical diagnosis or treatment [24–27], the present study screened the panel of serum metabolites in a model of AMI to diagnose human post-mortem cases with AMI type II, and these subjects could have nonetheless experienced a sudden acute ante-mortem ischemia just prior to death in addition to having occlusive coronary atherosclerotic plaques.

In the present study, a total of 28 endogenous metabolites were identified using untargeted metabolomics in rat serum after the AMI event. These metabolites were mainly involved in amino acids, bile acids, glucose, fatty acids, etc. However, although the types of the 28 differential metabolites are roughly the same with prior AMI metabolomics studies [19–21, 24–27], there were fewer identical metabolites in most of these studies. The variation in metabolites was mainly due to limitations of current analytical approaches in metabolomics, such as the bias of different instrument platform and lacking a systematic method for metabolite extraction [28, 29], which indicated that a standard method for metabolomic analysis needs to be developed which could be applied in different laboratories. In addition, the variation may also be caused by the different experimental animal models and species differences, including rats, dogs, pigs, and humans with coronary disease [20, 24–27].

As described previously, the diagnosis of cause in the first 6 h after onset of myocardial ischemia is a great challenge. 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 that represent potent tools for metabolomics analysis have been increasingly applied to the classification and data mining of complex UPLC-HRMS data [30]. 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 [31, 32]. And some reports proved that machine learning algorithms such as support vector machine, genetic algorithm, and random forest could be used to classify metabolomic data analysis and provide a high accuracy rate in diagnosing disease [31–33].

**Fig. 5** Boxplots of the 7 potential biomarkers showing a significant difference level between the AMI and other rat groups. a L-threonic acid, b N-acetyl-L-cysteine, c CMP, d glycocholic acid, e L-tyrosine, f cholic acid, and g glycourso deoxycholic acid. * represents $P<0.05$ compared with AMI group, ** represents $P<0.01$ compared with AMI group, *** represents $P<0.001$ compared with AMI group.

**Fig. 6** Validation of the MLP classification model for AMI diagnosis in rat and human samples. a Receiver operator characteristic (ROC) curve of MLP predicting model in rat samples. b Confusion matrix of AMI and control groups in rat. c ROC curve of MLP predicting model in human samples. d Confusion matrix of AMI and control groups in human samples.
In this experiment, we investigated the accuracy and efficacy of different metabolite combinations in inferring AMI using 5 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 could be improved by comparing different small molecule combinations of metabolites and employing appropriate machine learning algorithms (Fig. 4). To validate the model, MLP model was used to classify the 17 forensic autopsy-based blood samples and resulted in accuracy of 88.23% and ROC of 0.89. These results demonstrated that the 7 metabolites might be the biomarkers for clinical diagnosis or post-mortem identification. To the best of our knowledge, this is the first study to assess the capability of classification model based on untargeted metabolomics from well-controlled animal model to predict cause of death in autopsy 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 [34]. Zhang BC et al. found that higher serum total bile acid level was an independent predictor of high-risk coronary plaques in asymptomatic individuals [35], which indicated that serum bile acid level might forecast the AMI and the occurrence of SCD. Among the 7 essential metabolites in our study, there are 3 bile acids (cholic acid, glycocholic acid, and glycoursoheoxycholic acid) changes in serum of AMI. These results showed that the occurrence of myocardial ischemia leads to abnormal lipid metabolism, especially bile acid biosynthesis and metabolism. Sun L et al. found that the level of threonic acid was markedly down-regulated in the isoproterenol (ISO)-induced AMI rats and it was considered one of the most associated cecal metabolites with AMI [36]. Nevertheless, the serum level of threonic acid in this study 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. The “metabolic shift” could be the molecular autopsy indicator for medicolegal investigation in SCD cases. Our research provides a basis for the translational research from bench to daily practice for the diagnosis of acute myocardial ischemia. In the future, the standardization methods of metabolomics applied for cadaveric samples should be studied and the validation of diagnostic models based on the machine learning algorithms for SCD under different causes of death by increasing the sample size of humans.

Conclusions

In summary, there were 28 different metabolites identified by the metabolomics analysis from AMI rats sera. A panel of 7 valuable metabolites among 28 was screened by multiple machine learning algorithms and validated in autopsy cases. The constructed MLP classification model based on metabolomics data has a high diagnostic performance for AMI rats and the forensic autopsy-based blood samples. Therefore, combining the metabolomics approach and machine learning algorithms to establish diagnostic models has been demonstrated as a potentially helpful strategy for diagnosing AMI and SCD.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00414-022-02816-y.

Funding This research was supported by the Fund from Shanghai Key Laboratory of Forensic Medicine (Academy of Forensic Science) (Grant Numbers KF1803), the National Natural Science Foundation for Young Scientists of China (81901924), and the Natural Science Foundation for Young Scientists of Shanxi Province, China (201901D211334).

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

Declarations

Ethics approval and consent to participate The study was approved by the ethics committee of Shanxi Medical University, Shanxi, China (permit number: 2018LL346). Written informed consent statements were acquired from family of the deceased.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

1. Priori SG, Blomstrom-Lundqvist C, Mazzanti A et al (2015) 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 36:2793–2867
2. Hayashi M, Shimizu W, Albert CM (2015) The spectrum of epidemiology underlying sudden cardiac death. Circ Res 116:1887–1906
3. Campuzano O, Allegue C, Partemi S, Lglesias A, Oliva A, Brugada R (2014) Negative autopsy and sudden cardiac death. Int J Legal Med 128:599–606
4. Visona SD, Benati D, Monti MC, Galie M, Andreollo L, Frontini A, Ocsculati A (2018) Diagnosis of sudden cardiac death due to early
myocardial ischemia: an ultrastructural and immunohistochemical study. Eur J Histochem 62:2866.
5. Aljakna A, Fracasso T, Sabatasso S (2018) Molecular tissue changes in early myocardial ischemia: from pathophysiology to the identification of new diagnostic markers. Int J Legal Med 132:425–438.
6. Michaud K, Basso C, d’Amati G, Giordano C, Kholová I, Preston SD, Rizzo S, Sabatasso S, Sheppard MN, Vink A, van der Wal AC, Association for European Cardiovascular Pathology (AECVP) (2020) Diagnosis of myocardial infarction at autopsy: AECVP reappraisal in the light of the current clinical classification. Virchows Arch 476:179–194.
7. Rahimi R, Dahili ND, Anuar ZK, Anuar Zainun K, Mohd Kasim NA, Md Noor S (2018) Post mortem troponin T analysis in sudden death: Is it useful? Malays J Pathol 40:143–148.
8. Beausire T, Faouzi M, Palmiere C, Fracasso T, Michaud K (2018) High-sensitive cardiac troponin hs-TnT levels in sudden deaths related to atherosclerotic coronary artery disease. Forensic Sci Int 289:238–243.
9. Brion M, Sobrino B, Martinez M, Blanco-Vere A, Carracedo A (2015) Massive parallel sequencing applied to the molecular autopsy in sudden cardiac death in the young. Forensic Sci Int Genet 18:160–170.
10. Wu J, Wu Q, Dai W, Kong J, Lv J, Yu X, Wang X, Wang D (2018) 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 132:439–448.
11. Santori M, Blanco-Vere A, Gil R et al (2015) Bread-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 100:952–956.
12. Michaud K, Lesta MM, Fellmann F, Mangin P (2008) Molecular autopsy of sudden cardiac death: from post-mortem to clinical approach. Rev Med Suisse 4:1590–1593.
13. Johnson CH, Ivanisevic J, Sziudzag K (2016) Metabolomics: beyond biomarkers and towards mechanisms. Nat Rev Mol Cell Biol 17:451–459.
14. Cao J, Jin QQ, Wang GM et al (2018) Comparison of the serum metabolite signatures based on (1)H NMR between patients and a rat model of deep vein thrombosis. Sci Rep 8:7837.
15. Yu Z, Zhang H, Yu M, Ye Q (2015) Analysis of gene expression during the development of congestive heart failure after myocardial infarction in rat models. Int Heart J 56:444–449.
16. McGarrah RW, Crown SB, Zhang GF, Shah SH, Newgard CB (2018) Cardiovascular metabolomics. Circ Res 122:1238–1258.
17. Au A (2018) Metabolomics and lipidomics of ischemic stroke. Adv Clin Chem 85:31–69.
18. Li Y, Zhang D, He Y et al (2017) Investigation of novel metabolites potentially involved in the pathogenesis of coronary heart disease using a UHPLC-QTOF/MS-based metabolomics approach. Sci Rep 7:15357.
19. Khan A, Choi Y, Back JH, Lee S, Lee SH, Park YH (2019) High-resolution metabolomics study revealing l-homocysteine sulfenic acid, cysteic acid, and carnitine as novel biomarkers for high acute myocardial infarction risk. Metabolism 104:154051.
20. Wang X, Wang D, Wu J, Yu X, Lv J, Kong J, Zhu G, Su R (2017) Metabolic characterization of myocardial infarction using GC-MS-based tissue metabolomics. Int Heart J 58:441–446.
21. Gao Y, Gao J, Chen C, Wang H, Guo J, Wu R (2015) Cardioprotective effect of poydatin on ventricular remodeling after myocardial infarction in coronary artery ligation rats. Planta Med 81:568–577.
22. Mamoshina P, Volosnikova M, Ozernov IV, Putin E, Skibina E, Cortese F, Zhavoronkov A (2018) Machine learning on human muscle transcriptomic data for biomarker discovery and tissue-specific drug target identification. Front Genet 9:242.
23. Niu J, An G, Gu Z, Li P, Liu Q, Bai R, Sun J, Du Q (2020) Analysis of sensitivity and specificity precise recognition of neutrophils during regeneration of contused skeletal muscle in rats. Forensic Sci Res. https://doi.org/10.1080/20961790.2020.1713432.
24. van der Vusse GJ, Glazt JF, Stam HC, Reneman RS (1992) Fatty acid homeostasis in the normoxic and ischemic heart. Physiol Rev 72:881–940.
25. Cechova R, Tran T, Cechova A, Testerink B, Vorselen D, Koudela B, van Beuscho C, Vreeburg M, Willey L, Greiner S, Maguire S, Berntson G, de Jong M, Veening S, de Vos M, B-net Group, van Dijk W, van Dongen J, van der Vlist M (2020) Novel machine learning algorithm for arrhythmia prediction. Preprint at bioRxiv: 2020.07.03.220408.
26. Ashrafian H, Neubauer S (2009) Metabolomic profiling of cardiac substrate utilization: fanning the flames of systems biology? Circulation 119:1700–1702.
27. Shimokawa J, Taylor TG, Venable PW, Rhodes NL, Gil RB, Warren M, Wende AR, Abel ED, Cox J, Spitzer KW, Zaitsev AV (2013) Metabolic determinants of electrical failure in ex-vivo canine model of cardiac arrest: evidence for the protective role of inorganic pyrophosphate. PLoS One 8:e57821.
28. Schirrme-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA (2016) Untargeted metabolomics strategies-challenges and emerging directions. J Am Soc Mass Spectrom 27:1897–1905.
29. Alakwaa FM, Chaudhary K, Garmire LX (2018) Deep learning accurately predicts estrogen receptor status in breast cancer metabolomics data. J Proteome Res 17:337–347.
30. Wang X, Chen H, Chang C, Jiang M, Wang X, Xu L (2017) 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 Sci Res. https://doi.org/10.1016/j.chroma.2017.08.008.
31. Jiang M, Liang Y, Pei Z, Wang X, Zhou F, Wei C, Feng X (2019) Diagnosis of breast hyperplasia and evaluation of RuXian-I based on metabolomics deep belief networks. Int J Mol Sci 20:2620.
32. Alakwaa FM, Chaudhary K, Garmire LX (2018) Deep learning and early childhood. Arch Dis Child 100:952–956.
33. van der Vusse GJ, Glazt JF, Stam HC, Reneman RS (1992) Fatty acid homeostasis in the normoxic and ischemic heart. Physiol Rev 72:881–940.
34. Ashrafian H, Neubauer S (2009) Metabolomic profiling of cardiac substrate utilization: fanning the flames of systems biology? Circulation 119:1700–1702.
35. Zhang BC, Chen JH, Xiang CH, Su MY, Zhang XS, Ma YF (2019) Inorganic pyrophosphate. PLoS One 8:e57821.
36. Schirrme-Rutledge AC, Codreanu SG, Sherrod SD, McLean JA (2016) Untargeted metabolomics strategies-challenges and emerging directions. J Am Soc Mass Spectrom 27:1897–1905.
37. Alakwaa FM, Chaudhary K, Garmire LX (2018) Deep learning and early childhood. Arch Dis Child 100:952–956.