



OPEN ACCESS

EDITED BY

Angela Di Somma,
University of Naples Federico II, Italy

REVIEWED BY

Gloria Riitano,
Sapienza University of Rome, Italy
Manish Kumar Sinha,
Vanderbilt University, United States
Clarissa Braccia,
San Raffaele Hospital (IRCCS), Italy

*CORRESPONDENCE

Jingchun Song,
✉ songjingchun@126.com

[†]These authors have contributed equally
to this work

RECEIVED 23 April 2025

ACCEPTED 11 June 2025

PUBLISHED 20 June 2025

CITATION

Zeng Q, Lin Q, He L, Zhong L, Zhou Y, Deng X,
Zhang N, Song Q and Song J (2025)
Developing a molecular diagnostic model for
heatstroke-induced coagulopathy: a
proteomics and metabolomics approach.
Front. Mol. Biosci. 12:1616073.
doi: 10.3389/fmolb.2025.1616073

COPYRIGHT

© 2025 Zeng, Lin, He, Zhong, Zhou, Deng,
Zhang, Song and Song. This is an open-access
article distributed under the terms of the
[Creative Commons Attribution License \(CC
BY\)](#). The use, distribution or reproduction in
other forums is permitted, provided the
original author(s) and the copyright owner(s)
are credited and that the original publication
in this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Developing a molecular diagnostic model for heatstroke-induced coagulopathy: a proteomics and metabolomics approach

Qingbo Zeng^{1†}, Qingwei Lin^{1†}, Longping He¹, Lincui Zhong¹,
Ye Zhou¹, Xingping Deng¹, Nianqing Zhang², Qing Song^{3,4} and
Jingchun Song^{1,4*}

¹Intensive Care Unit, The 908th Hospital of Chinese PLA Logistic Support Force, Nanchang, China,

²Intensive Care Unit, Nanchang Hongdu Hospital of Traditional Chinese Medicine, Nanchang, China,

³Department of Critical Care Medicine, Hainan Hospital, Chinese PLA General Hospital, Sanya, China,

⁴Scientific Research Department, Heatstroke Treatment and Research Center of PLA, Sanya, China

Background: Heatstroke (HS) is becoming more concerning, with coagulopathy contributing to higher mortality. The aim of this study was to analyze the metabolomic and proteomic profiles associated with heatstroke-induced coagulopathy (HSIC) and to develop a molecular diagnostic model based on proteomic and metabolomic patterns.

Methods: This study included 41 HS patients from the Department of Critical Care Medicine at a comprehensive teaching hospital. Plasma proteins and metabolites from HSIC and non-heatstroke-induced coagulopathy (NHSIC) patients were compared using LC-MS/MS. Multivariate and univariate statistical analyses identified differentially expressed proteins (DEPs) and metabolites (DEMs). Functional annotation and pathway enrichment analyses were performed using the GO and KEGG databases, and machine learning models were developed using candidate proteins selected by LASSO and Boruta algorithms to diagnose HSIC. Finally, bioinformatic analysis was used to integrate the results of proteomics and metabolomics to find the potential mechanisms of HSIC.

Results: A total of 41 patients participated in the study, with 11 cases in the HSIC group and 30 cases in the NHSIC group. Significant differences were observed between the groups in temperature, heart rate, white blood cell count, platelet count, liver function, coagulation markers, APACHE II score, and GCS score. Survival analysis revealed that the heatstroke group had a higher mortality risk. A total of 125 DEPs and 110 DEMs were identified, primarily enriched in energy regulation-related pathways and lipid and carbohydrate metabolism. Additionally, three optimal predictive models (AUC >0.9) were developed and validated for classifying HSIC from HS individuals based on proteomic patterns and machine learning, with the logistic regression model showing the best diagnostic performance (AUC = 0.979, sensitivity = 81.8%, specificity = 96.7%), highlighting lactate dehydrogenase A chain (LDHA), neutrophil gelatinase-associated

lipocalin (NGAL), prothrombin and glucan-branching enzyme (GBE) as key predictors of HSIC.

Conclusion: The study uncovered critical metabolic and protein changes linked to heatstroke, highlighting the involvement of energy regulation, lipid metabolism, and carbohydrate metabolism. Building on these findings, an optimal machine learning diagnostic model was developed to boost the accuracy of HSIC diagnosis, integrating LDHA, NGAL, prothrombin, and GBE as key biomarkers.

KEYWORDS

heatstroke, coagulopathy, machine learning, metabolomics, proteomics

1 Introduction

HS is a severe condition marked by a rapid increase in body temperature above 40°C and central nervous system dysfunction, leading to inflammation, organ failure, and potentially death (Bouchama and Knochel, 2002; Liu et al., 2020a). It can be exertional, due to intense activity, or classic, affecting vulnerable populations like the elderly (Bouchama et al., 2022). The overall incidence of heatstroke in 2020 was 0.36 per 1,000 United States military soldiers/years (Author Anonymous, 2021), while in Japan, it ranged from 37.5 per 100,000 (95% CI, 36.8–38.2) between 2015 and 2017 to 74.4 (95% CI, 72.7–76.1) in 2018 (Ogata et al., 2021). HS is frequently accompanied by coagulation disorders and coagulopathy (Iba et al., 2022; Huisse et al., 2008). In this case, both coagulation abnormalities and HS act as contributing factors. This will worsen the patient's condition, increase the risk of disseminated intravascular coagulation (DIC), and make treatment both challenging and less effective. Therefore, there is an urgent need to explore the risk factors and elucidating the mechanisms associated with the early diagnosis, classification of risks, and prediction of outcomes between HS and HSIC underlying their clinical development. Early identification is crucial as it facilitates prompt therapeutic intervention, potentially preventing disease progression to life-threatening complications.

Currently, we lack precise indicators and recognized diagnostic criteria to identify high-risk HSIC populations, hindering early interventions. This is mainly due to an incomplete understanding of HS's pathogenesis and the specificity of HSIC. Researchers have analyzed the link between HS and coagulopathy, focusing on blood cells' role in inflammation and coagulation, especially platelet dysfunction (Iba et al., 2025). Additionally, researchers found severe endothelial damage and abnormal coagulation activation during HS, and identified key coagulation and endothelial risk factors related to DIC to establish diagnostic models (Liu et al., 2025; Zeng et al., 2023). The advancement of omics technologies, such as metabolomics and proteomics, provide a novel approach for diagnosing HSIC and deepen our understanding of its occurrence and progression from an immune-metabolism perspective.

In this study, we applied metabolomics and quantitative proteomics to quantitatively analyze the plasma of patients with HSIC and NHSIC (Figure 1). We identified differentially expressed metabolites (DEMs), DEPs, and key signaling pathways.

Additionally, we developed diagnostic models based on proteomic patterns and machine learning. The performance of these models was evaluated using metrics such as accuracy, sensitivity, specificity, and area under the curve (AUC) to determine their potential for early detection of HSIC in clinical settings.

2 Materials and methods

2.1 Study population

This study included 41 HS patients enrolled from the Department of critical care medicine of the 908th Hospital of PLA joint logistic support force, from June 2022 to February 2024. The patients were divided into two groups: 11 cases in the HSIC group and 30 cases in the NHSIC group. The study was conducted in accordance with the ethical guidelines outlined in the Declaration of Helsinki, and informed consent was obtained from all participants or their family members. The study was approved by the ethics committee of our hospital.

Patients were diagnosed with HS or HSIC based on the Bouchama's HS criteria (Bouchama and Knochel, 2002) and a newly proposed HSIC score (Lin et al., 2024). HS was defined by meeting the following criteria: (1) medical history of exposure to high temperature, high humidity or high-intensity exercise; (2) core temperature over 40°C; (3) CNS-related changes, including coma, convulsions, delirium, or abnormal behavior. HSIC scores were calculated according to as follows: Core temperature <40°C was scored as 0 points, ≥40°C but <42°C as 1 point, ≥42°C as 2 points; D-dimer of 2-fold increase over baseline was scored as 0 point, ≥ 2-fold but < 5-fold increase over baseline as 1 point, ≥ 5-fold increase over baseline as 2 points; prothrombin time (PT) prolongation of <2 s was scored as 0 point, ≥2 s but <4 s as 1 point, ≥4 s as 2 points. HSIC was diagnosed as a total score at least 3. Patients were excluded from our study if they were younger than 18 years old, they had a congenital coagulation disorder or chronic disease of the liver, or they were using anticoagulant drugs at the time of enrollment.

2.2 Sample collection process

Venous blood samples were drawn in the early morning after a minimum of 8 h of fasting and placed into a dry blood collection

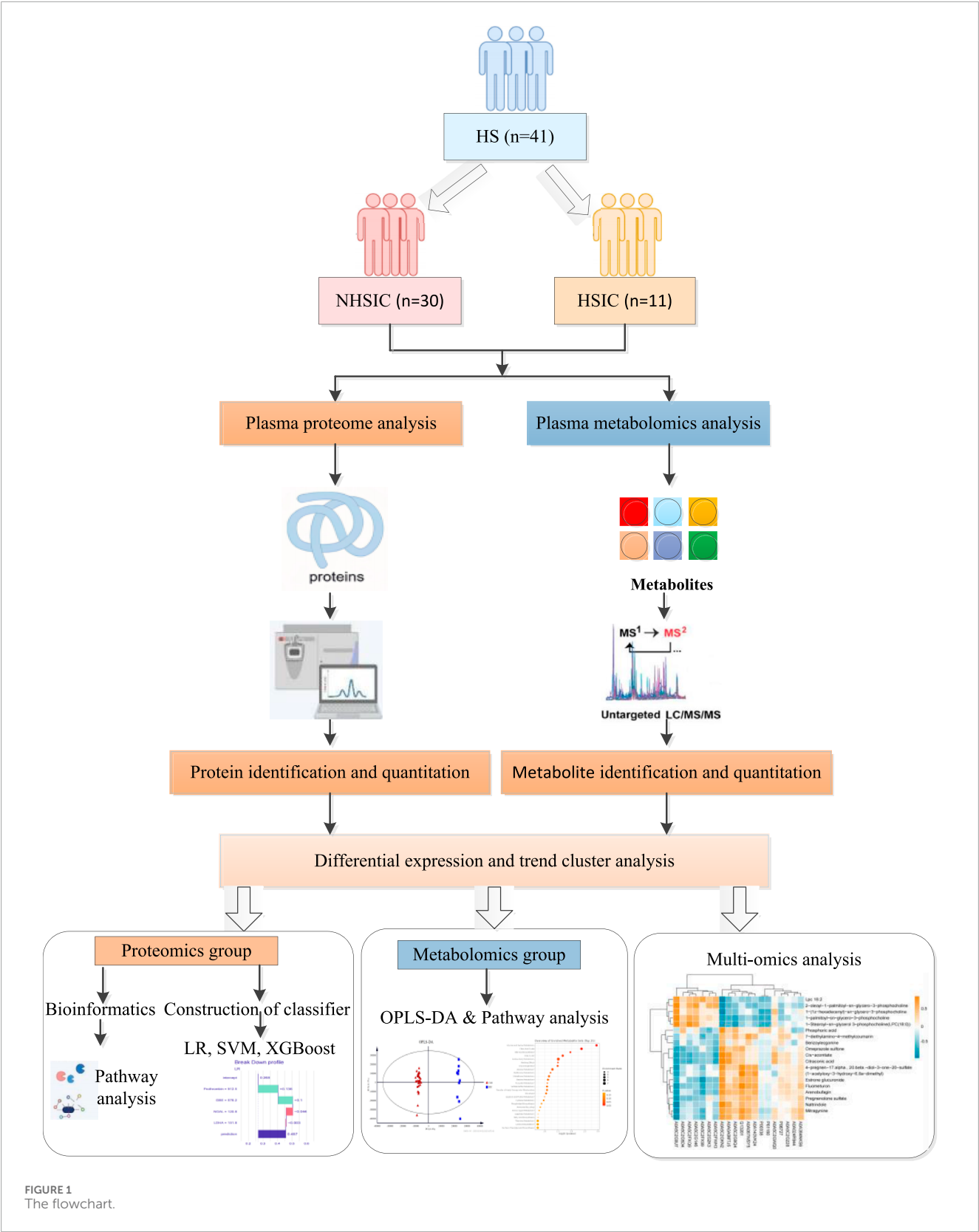


FIGURE 1
The flowchart.

tube. The samples were then centrifuged at 4244 g for 10 min at 4°C. The serum obtained was separated and stored at −80°C.

2.3 Proteomics

In our study, we analyzed serum samples from 41 participants using DIA quantitative proteomics. The total protein from each sample was extracted, with a portion used for protein concentration determination and SDS-PAGE, while the remainder underwent trypsin digestion. After desalting, LC-MS/MS was employed to identify the peptides present. LC-MS/MS analysis was conducted on a timsTOF Pro 2 system (Bruker) coupled with Evosep One LC, employing both DDA (scan range m/z 100–1,700, ion mobility 0.75–1.35 Vs/cm², 8 PASEF MS/MS cycles) for spectral library generation and 4-window DIA for quantitative profiling. Spectronaut pulsar software was used to comprehensively search the raw data and compare it against the known UniProtKB human proteome database. The false discovery rate (FDR) for both DDA and DIA data was set at 0.01. DEPs were defined with a p -value <0.05 and $|\log_2(\text{FC})| > 1$.

GO and KEGG pathway analyses were used to assess protein functions. In the GO annotation process, the protein sequences of the differentially expressed proteins were first searched locally using the NCBI BLAST + client software (ncbi-blast-2.2.28+win32.exe) and InterProScan to identify homologous sequences. Gene ontology (GO) terms were then mapped, and the sequences were annotated using Blast2GO software, with the results visualized using R scripts. For KEGG annotation, the annotated proteins were blasted against the Kyoto Encyclopedia of Genes and Genomes (KEGG) database (<http://geneontology.org/>) to obtain their KEGG orthology identifications, which were subsequently mapped to relevant pathways in KEGG. A p -value less than 0.05 in the pathway enrichment test, along with a protein count greater than 5, were used as criteria for determining significance.

2.4 Metabolomics

Serum samples from 41 subjects were analyzed using metabolomics. Plasma samples were thawed at 4°C, and 100 μL of each sample was mixed with 400 μL of pre-cooled methanol/acetonitrile (1:1, v/v) to precipitate proteins. The mixture was centrifuged at 14,000 g and 4°C for 20 min. The supernatant was dried under vacuum centrifugation and re-dissolved in 100 μL of acetonitrile/water (1:1, v/v). After centrifugation at 14,000 g and 4°C for 15 min, the supernatant was injected for analysis. An UHPLC system (1,290 Infinity LC, Agilent Technologies) coupled to a quadrupole time-of-flight mass spectrometer (AB Sciex TripleTOF 6600) was used. Hydrophilic interaction chromatography (HILIC) was performed using a 2.1 mm $\times$ 100 mm ACQUITY UPLC BEH Amide 1.7 μm column (Waters, Ireland). In both ESI positive and negative modes, mobile phases consisted of A = 25 mM ammonium acetate and 25 mM ammonium hydroxide in water, and B = acetonitrile. The gradient started at 95% B for 0.5 min, decreased to 65% in 6.5 min, to 40% in 1 min (held for 1 min), then returned to 95% in 0.1 min with a 3 min re-equilibration. Source parameters: Gas1 = 60, Gas2 = 60, CUR = 30, temperature = 600°C,

ISVF = ± 5500 V. MS range: 60–1,000 Da (0.20 s/spectra); MS/MS: 25–1,000 Da (0.05 s/spectra), CE = 35 V (± 15 eV), DP = ± 60 V, IDA with 10 candidate ions per cycle. An UHPLC (Vanquish, Thermo) coupled to an Orbitrap (Q Exactive HF-X/Q Exactive HF) was also employed. The same HILIC column and mobile phases were used. Gradient: 98% B for 1.5 min, decreased to 2% in 10.5 min, held for 2 min, then returned to 98% in 0.1 min with a 3 min re-equilibration. MS range: 80–1,200 Da, resolution 60,000 (100 ms); MS/MS: 70–1,200 Da, resolution 30,000 (50 ms), exclude time: 4 s. Raw MS data were converted to MzXML files using ProteoWizard MSConvert, imported into XCMS for peak picking (centWave m/z = 10 ppm, peakwidth = c (10, 60)), and annotated with CAMERA. Compound identification was based on m/z and MS/MS comparison with an in-house database. The positive and negative data were merged to create a combined dataset, which was then imported into R software. After sum-normalization, the data were analyzed using the R package (ropis), performing multivariate analysis including PCA and OPLS-DA with Pareto scaling. 7-fold cross-validation and 200 response permutation testing assessed model robustness. VIP values were calculated to determine each variable's contribution to classification. Student's t -test was used to evaluate differences between two groups, with VIP >1 and $p < 0.05$ for screening significant metabolites.

2.5 Statistical analyses

All statistical analyses were performed using R software (version 4.1.2). Continuous variables were presented as means $\pm$ SD or medians with IQR, depending on distribution. Differences between groups were assessed using t -tests for data following a normal distribution or Mann-Whitney U tests for data following a non-normal distribution. Categorical variables were compared using chi-square or Fisher's exact tests. A p -value <0.05 was considered significant. Feature selection was conducted with LASSO and Boruta algorithms. Predictive models were developed using LR, XGBoost, and SVM. Model performance was evaluated with metrics such as accuracy, sensitivity, specificity, AUC, kappa, and F1 score. ROC and PR curves were used to assess discriminative power, while DCA validated clinical utility. DALEX was used for model interpretability.

3 Results

3.1 Participant characteristics

A total of 41 patients with HS were enrolled in this study after screened for eligibility. All of patients with HS were divided into HSIC group ($n = 11$, 9 males and 2 females), and NHSIC group ($n = 30$, 25 males and 5 females). Inter-group comparisons revealed statistically significant differences in temperature, heart rate, white blood cells (WBC), platelets (PLT), aspartate transaminase (AST), ALT (alanine transaminase), total bilirubin (TBIL), creatinine (Cr), lactate (Lac), creatine kinase (CK), MB(myoglobin), PT, INR (international normalized ratio), APTT (activated partial thrombin time), TT (thrombin time), d-dimer, APACHE II(Acute Physiology and Chronic Health Evaluation II) score, GCS(Glasgow Coma Scale) score and MODS (multiple organ dysfunction

syndrome). While age, gender, risk of coronary disease and diabetes, probability of hypertension, Hemoglobin, HCT (hematocrit), and FIB (fibrinogen) showed no statistical differences between the two groups. The comparisons between HSIC group and NHSIC group are depicted in [Table 1](#).

3.2 APACHE II score and survival analysis in the study groups

Compared with NHSIC group, the APACHE II score in HSIC group is significantly increased ([Table 1](#); [Figure 2A](#)). The cumulative risk curve of adverse outcome according to the study groups is shown in [Figure 2B](#). The result demonstrated that individuals in NHSIC group were at the higher risk of death compared with HSIC group.

3.3 Proteins identification and differential proteins analysis

We identified a total of 1907 proteins. For differential expression analysis, p value <0.05 were considered statistically significant and $|\log_2(\text{FC})| >1$ was considered up- or downregulated, respectively. A total of 125 DEPs were screened in the NHSIC group vs. HSIC group. A comparison of NHSIC and HSIC patients found that there were 6 and 119 upregulated and downregulated proteins, respectively ([Figure 3A](#)). Hierarchical clustering algorithm was used to perform cluster analysis on the DEPs of NHSIC and HSIC groups, and the data was displayed in the form of heat map ([Figure 3B](#)).

3.4 Functional analysis of differentially expressed proteins

We performed GO functional annotation on the differentially expressed proteins (DEPs) ([Figure 4A](#)). In the NHSIC vs. HSIC group, the difference in protein expression was greatest in Biological Process (BP), Molecular Function (MF), and Cellular Component (CC). Concentrated results were in negative regulation of coagulation, lipoprotein particle receptor binding, and lipoprotein particle. We used KEGG to analyze the signaling pathways of 125 DEPs in the NHSIC vs. HSIC group that were mainly concentrated in complement and coagulation cascades, cholesterol metabolism, and neuroactive ligand–receptor interaction ([Figure 4B](#)). From the 125 DEPs in the NHSIC vs. HSIC group, 36 proteins were involved in protein interactions. Furthermore, PPIs analysis showed that APOH and PLG had the highest degree scores in the network ([Figure 4C](#)), suggesting that they play a role in the pathological process of HS.

3.5 Identification and further analysis of candidate DEPs involved in HSIC

LASSO regression introduces a penalty function to continuously compress the coefficients, streamline the model, and avoid multicollinearity and overfitting, thereby achieving the effect of variable selection. When a standard error of the minimum distance

λ was 0.012, 27 feature coefficients were nonzero ([Figures 5A–C](#)). The process and results of feature selection using Random Forest are shown in [Figures 5D,E](#), which identified 16 important variables predisposing to HSIC in heatstroke patients. The two methods identified overlapping proteins, two upregulated (LDHA and NGAL), two downregulated (Prothrombin and GBE) ([Figure 5F](#)). Then, these four most vital variables were selected for further analysis. As the principal component analysis result in [Figure 5G](#), the four DEPs aforementioned can clearly distinguished NHSIC and HSIC, which indicated that they may play key roles in the diagnosis of HSIC. The correlations of the proteins were also analyzed as shown in [Figure 5H](#). The absence of significant correlations among these proteins indicates they do not exhibit functional similarities.

3.6 Construction and assessment of XGBoost, LR and SVM model for HSIC diagnosis

XGBoost, LR and SVM algorithms were used to construct models based on selected proteins. We used AUC, accuracy, no information rate, balanced accuracy, kappa, sensitivity, specificity, precision, and F1 scores to comprehensively evaluate the model's performance. XGBoost had the largest AUC (0.991) and precision (1.0), followed by LR (AUC: 0.979; precision: 0.90) and SVM (AUC: 0.976; precision: 0.750) ([Table 2](#)). [Figure 6A](#) described the ROC curves for the three models. The accuracy, kappa, sensitivity, Balanced accuracy and F1 scores of LR were higher than those of XGBoost and SVM, as shown in [Table 2](#). Therefore, The LR had better clinical utility compared to XGBoost and SVM.

The ROC curves and PR curves for the three models indicated that the LR was better than XGBoost and SVM in discrimination ([Figures 6A,B](#)). The calibration curve fit was good ([Figure 6C](#)), indicating that the model had good diagnostic performance. Furthermore, we found the net benefits (NB) of LR model was higher than 0, with the greater NB in DCA clinical evaluation ([Figure 6D](#)), indicating the importance of the LR model for HSIC diagnosis. The DALEX package was used for logistic regression analysis to further demonstrate the importance of the four proteins in the model and the descending order of importance of these features in the model was LDHA, NGAL, Prothrombin and GBE ([Figure 6E](#)). Based on the confusion matrix, the assay had a sensitivity, specificity, PPV, and NPV of 81.8%, 96.7%, 90.0%, and 93.6% respectively ([Figure 6F](#)).

3.7 Model explainability and visualization

We present one instances in which interpretability analyses of the model predictions. The ML model predicted a 45.7% risk of HSIC based on four critical predictors. We found prothrombin and GBE being the two contributors to the increased risk of HSIC, whereas LDHA and NGAL reduced the model's diagnosis of HSIC ([Figure 6G](#)). We developed the interface to facilitate the use of the model to explore the relative contribution of HSIC probability factors in ICU patients. In the diagnosis view, the system invokes a diagnosis model, and the LR model diagnoses the

TABLE 1 Baseline demographic.

Item	NHSIC (N = 30)	HSIC (N = 11)	p-value	q-value
Male, n (%)	25 (83)	9 (82)	>0.990	>0.990
age, yr	31 (27–36)	49 (40–58)	0.059	0.084
Temperature, °C	36.8 (36.5–37.0)	38.4 (37.7–39.2)	0.001	0.003
Heart rate, min ⁻¹	70 (57–80)	110 (90–124)	<0.001	<0.001
Hypertension, n (%)	0 (0)	2 (18)	0.067	0.091
Diabetes, n (%)	1 (3.3)	0 (0)	>0.990	>0.990
Coronary disease, n (%)	0 (0)	1 (9.1)	0.270	0.320
WBC, ×10 ¹² /L	9.0 (6.7–11.4)	14.9 (12.0–20.8)	0.009	0.015
HGB, g/L	129 (119–142)	132 (115–140)	0.660	0.710
HCT, %	39.9 (37.7–43.7)	40.5 (35.9–44.4)	0.630	0.700
PLT, × 10 ⁹ /L	231 (178–284)	54 (30–103)	<0.001	0.001
ALT, U/L	27.1 (16.7–38.1)	110.3 (34.1–955.1)	0.001	0.003
AST, U/L	28.8 (21.7–51.4)	204.7 (65.7–476.1)	<0.001	<0.001
TBIL, μmol/L	14.3 (10.0–18.9)	22.5 (17.2–49.2)	0.003	0.005
Cr, μmol/mL	83.0 (74.8–103.2)	184.7 (104.6–299.3)	0.002	0.005
Lac, mmol/L	0.9 (0.7–1.9)	2.7 (1.800–9.8)	0.004	0.008
CK, U/L	331.7 (160.6–699.9)	1,521.0 (242.7–8,947.7)	0.009	0.014
MB, U/L	80.2 (52.1–317.3)	745.5 (580.8–852.1)	<0.001	<0.001
PT, s	13.4 (12.3–14.4)	19.1 (17.6–30.9)	<0.001	<0.001
INR	1.12 (1.05–1.20)	1.58 (1.45–2.52)	<0.001	<0.001
APTT, s	29.4 (27.9–31.4)	53.4 (37.5–120.0)	<0.001	<0.001
TT, s	16.4 (15.5–17.6)	18.9 (16.7–27.4)	0.005	0.010
FIB, g/L	2.13 (1.85–2.49)	1.93 (1.21–2.66)	0.490	0.560
D-dimer, μg/L	0.24 (0.19–0.870)	14.80 (3.75–29.97)	<0.001	<0.001
APACHE II score	11 (7–15)	28 (21–39)	<0.001	<0.001
GCS	13 (11–15)	4 (3–8)	<0.001	<0.001
MODS, n (%)	4 (13)	8 (73)	<0.001	0.002
Mortality, n (%)	1 (3.3%)	3 (27%)	0.052	0.077

Values are n (%), mean ± standard deviation or median (interquartile range), unless otherwise noted.
Abbreviations: WBC, white blood cells; HGB, hemoglobin; HCT, hematocrit; ALT, alanine transaminase; AST, aspartate transaminase; Cr, creatinine; TBIL, total bilirubin; Lac, lactate; CK, creatine kinase; MB, myoglobin; PLT, platelets; PT, prothrombin time; INR, international normalized ratio; APTT, activated partial thrombin time; TT, thrombin time; FIB, fibrinogen; APACHE II, score, Acute Physiology and Chronic Health Evaluation II; GCS, glasgow coma scale; MODS, multiple organ dysfunction syndrome.

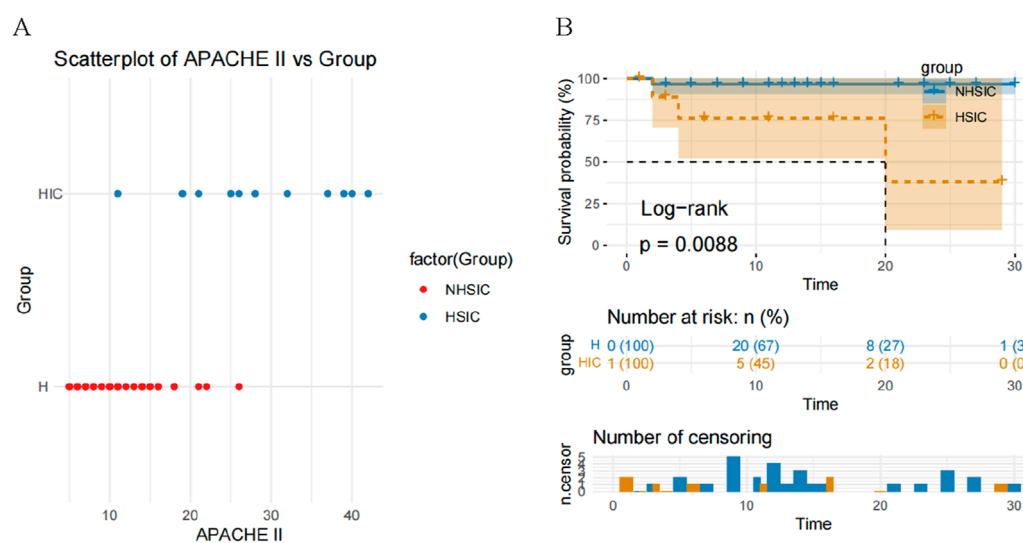


FIGURE 2

(A) A scatterplot of APACHE II score comparison in the NHSIC vs. HSIC group; (B) Kaplan-Meier curves for adverse outcome according to the study groups.

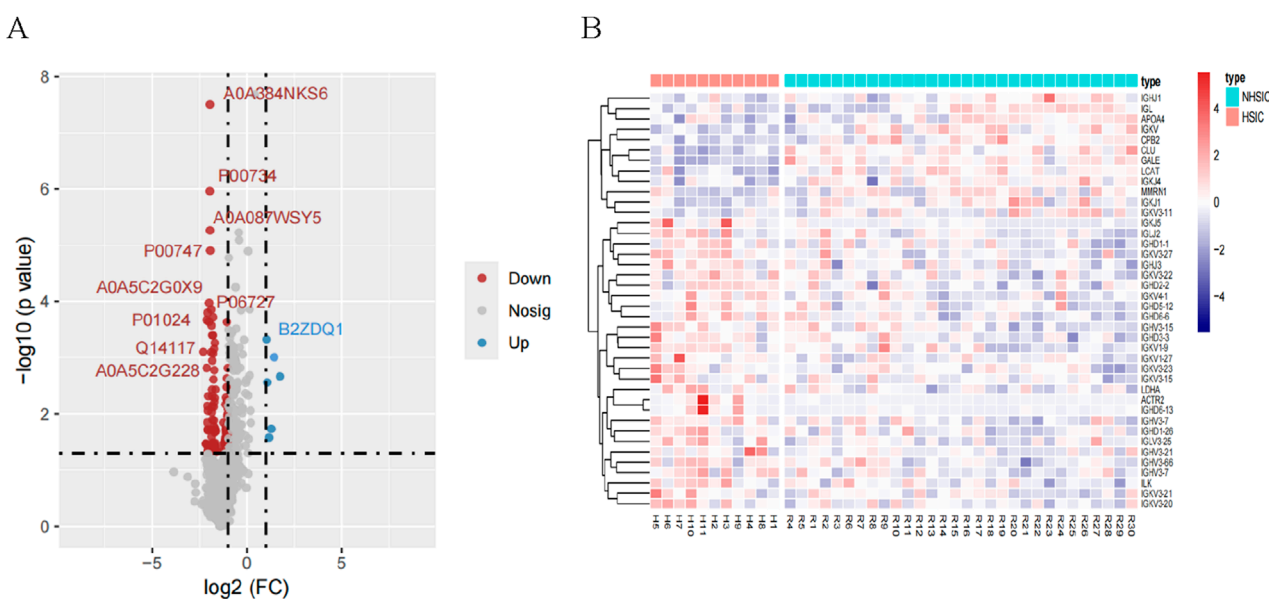


FIGURE 3

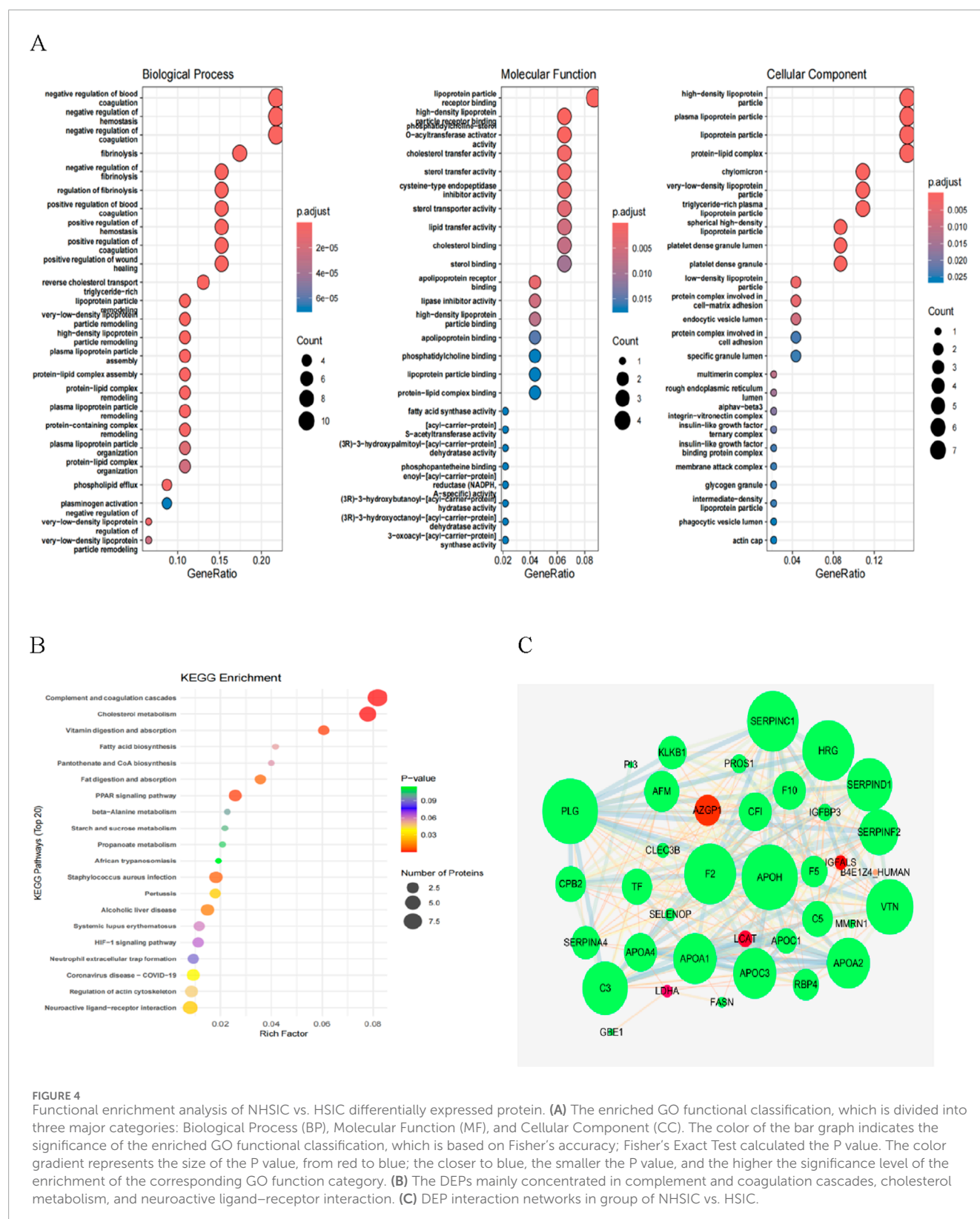
(A) A total of 125 differentially expressed proteins were screened in the NHSIC vs. HSIC group (6 upregulated and 119 downregulated). (B) Cluster analysis of DEPs in the group of NHSIC vs. HSIC.

patient's HSIC. The analysis results are visualized in a graphic view, which indicates the HIC probability of the patient input important features values (Figure 6H).

In addition, we also described the effect (positive or negative) of four proteins on the model. Figure 7 showed the relationship between LDHA, Prothrombin, NGAL, GBE and predicted HSIC. Higher LDHA and NGAL were associated with an increased risk of HSIC. Lower Prothrombin and GBE were associated with an increased risk of HSIC (Figure 7).

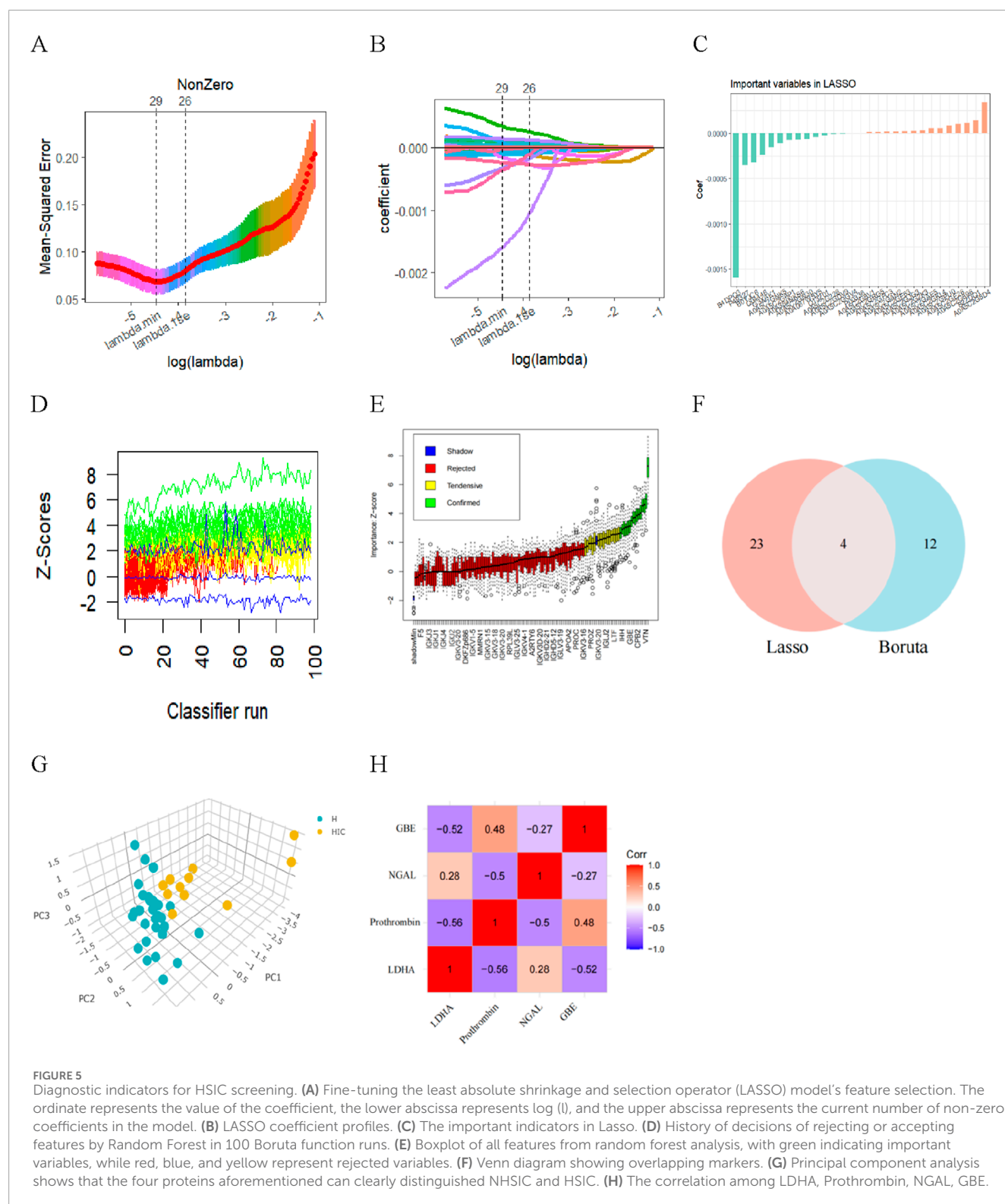
3.8 Metabolomic analysis

Figure 8A displays the results of principal component analysis (PCA), which clearly separates the samples into distinct groups. Figure 8B presents the OPLS-DA scores, highlighting a significant difference between the two groups. The robustness of the OPLS-DA model was validated through permutation analysis, yielding an R^2Y of 0.979 and a Q^2 value of 0.545. In Figure 8C, the negative Q^2 value confirms the absence of over-fitting, further



validating the model's reliability and effectiveness. Differential expression analysis, based on a Benjamini-Hochberg adjusted filter of <0.05 and a $\log_2$ fold change (FC) > 1.0 , revealed 110 significantly DEMs, including 30 upregulated and 80 downregulated

metabolites. These 110 DEMs are shown in the volcano plot in Figure 8D. Finally, Figure 8E illustrates the top 25 enriched metabolic pathways for these DEMs, as identified using the KEGG database.



3.9 Bioinformatic analysis-integrated analysis of proteomics and metabolomics

Through DIABLO algorithm, we identified distinct clustering patterns differentiating NHSIC and HSIC samples (Figure 9A). Subsequent integrative modeling of multi-omics features

revealed strong cross-correlations between proteomic and metabolomic profiles, highlighting their coordinated biological regulation (Figure 9B). To identify significant protein-metabolite interactions, we extracted the top 20 differentially expressed proteins (DEPs) and metabolites (DEMs) ranked by p-values and performed Pearson correlation analysis to evaluate their

TABLE 2 Performance of three machine learning models for predicting HSIC in critically ill patients.

Models	AUC	Accuracy	No information rate	Balanced accuracy	Kappa	Precision	F1 score	Sensitivity	Specificity
XGBoost	0.991	0.877	0.731	0.773	0.637	1.000	0.706	0.545	1.000
SVM	0.976	0.878	0.730	0.859	0.698	0.750	0.783	0.818	0.900
LR	0.979	0.927	0.732	0.892	0.808	0.900	0.857	0.818	0.967

AUC, area under curve; LR, logistic regression; XGBoost, eXtreme gradient boosting; SVM, support vector machine.

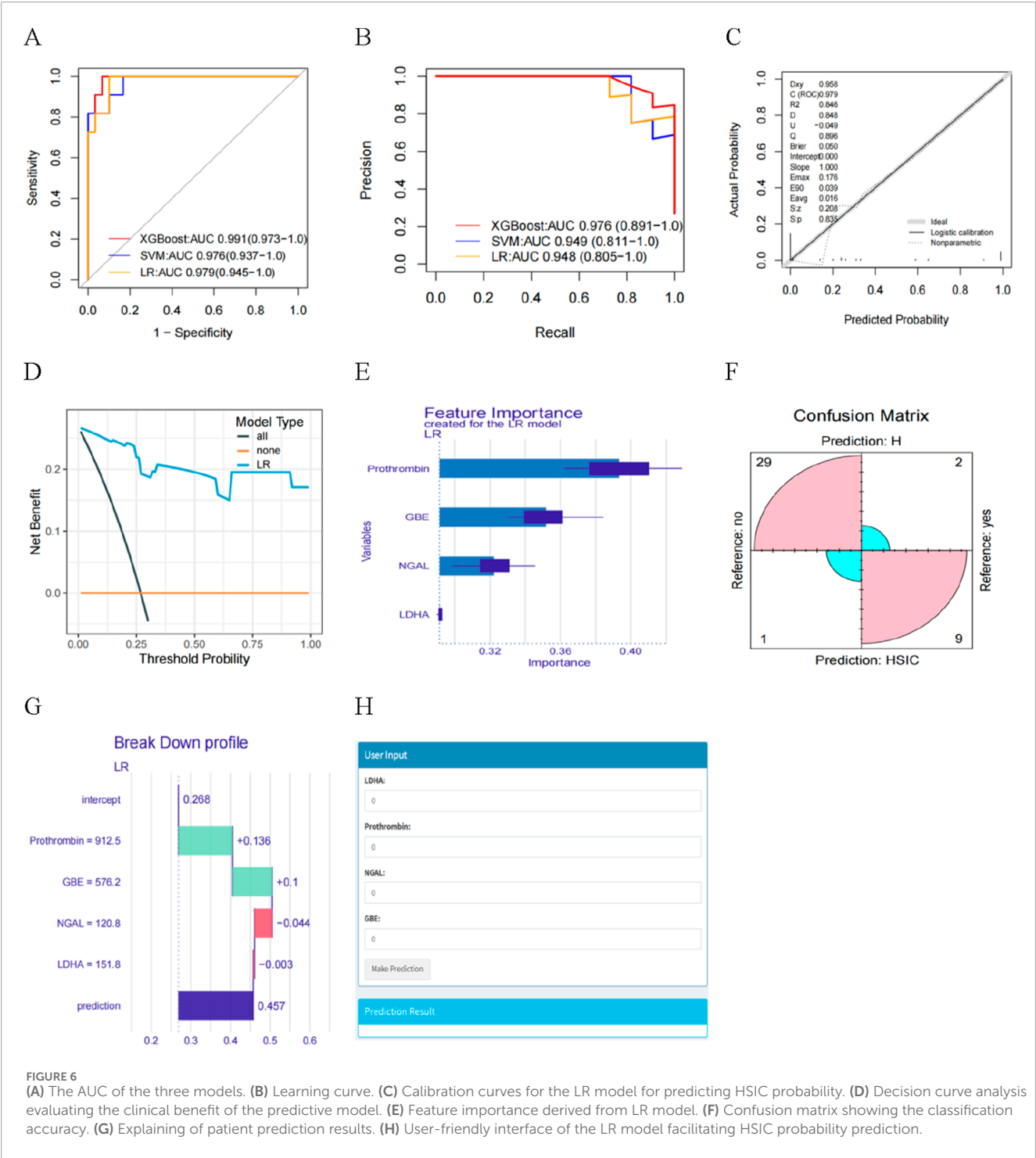
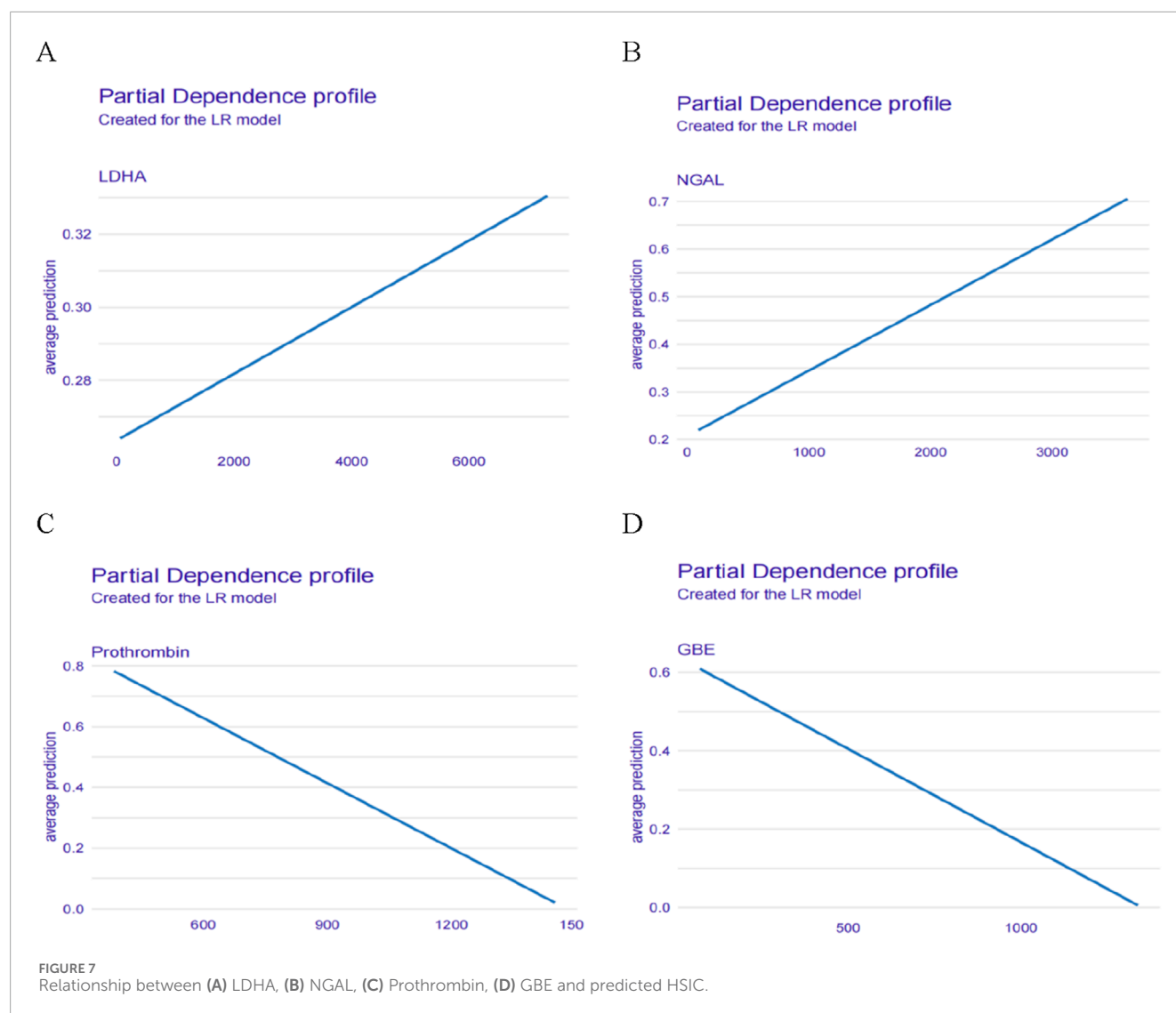


FIGURE 6 (A) The AUC of the three models. (B) Learning curve. (C) Calibration curves for the LR model for predicting HSIC probability. (D) Decision curve analysis evaluating the clinical benefit of the predictive model. (E) Feature importance derived from LR model. (F) Confusion matrix showing the classification accuracy. (G) Explaining of patient prediction results. (H) User-friendly interface of the LR model facilitating HSIC probability prediction.



associations, followed by the generation of a correlation heatmap (Figure 9D). Multi-omics clustering analysis revealed strong inter-dataset associations between features derived from the proteomic and metabolomic profiles (Figure 9D). Additionally, the differentially expressed proteins and metabolites were mapped to the KEGG database to assess the potential relationship. It was found that the AMPK signaling pathway, cholesterol metabolism, glycolysis/gluconeogenesis, neuroactive ligand-receptor interaction, propanoate metabolism, glucagon signaling pathway, glycerophospholipid metabolism, and pantothenate and CoA biosynthesis were significantly enriched (Figure 9E).

4 Discussion

This study explored the clinical characteristics, biomarkers, and the performance of predictive models for HSIC. It revealed significant differences in clinical markers such as APACHE II scores, liver function, and coagulation markers between the NHSIC and HSIC groups. There were 125 DEPs and 110 DEMs identified in the

comparison between the NHSIC and HSIC groups and Significant protein and metabolite alterations were found primarily in energy regulation and lipid metabolism pathways. Specific proteins, such as LDHA, NGAL, GBE, and prothrombin, were identified as potential biomarkers for HSIC diagnosis. Additionally, Machine learning models, including XGBoost, LR, and SVM, were developed for HSIC diagnosis, with the LR model showing the best diagnostic performance.

The clinical markers, including temperature, heart rate, AST, ALT, PT, and INR, observed in both the NHSIC and HSIC groups, provide valuable insights into the pathophysiology of HS and its complications. These variations align with previous studies showing that heatstroke causes organ dysfunction and coagulation abnormalities (He et al., 2022; Zhong et al., 2021). The higher APACHE II scores in the HSIC group further highlight the connection between the severity of heatstroke and poor prognosis, as high APACHE II scores are known to predict adverse outcomes (Shimazaki et al., 2020). The identification of 125 DEPs, reflecting changes in coagulation and lipoprotein metabolism, supports earlier studies (Li et al., 2023; Caputa et al., 2000). The functional analysis

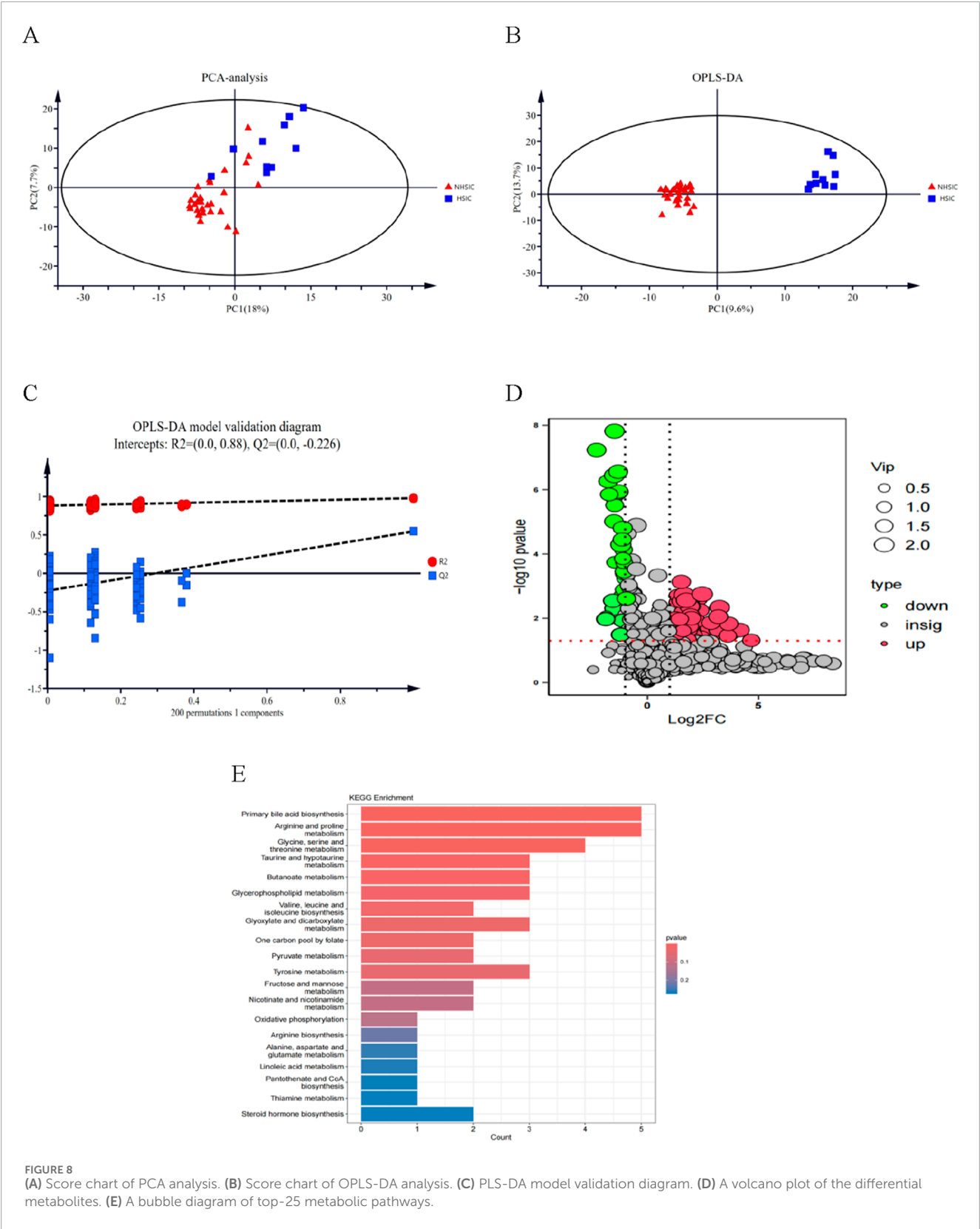
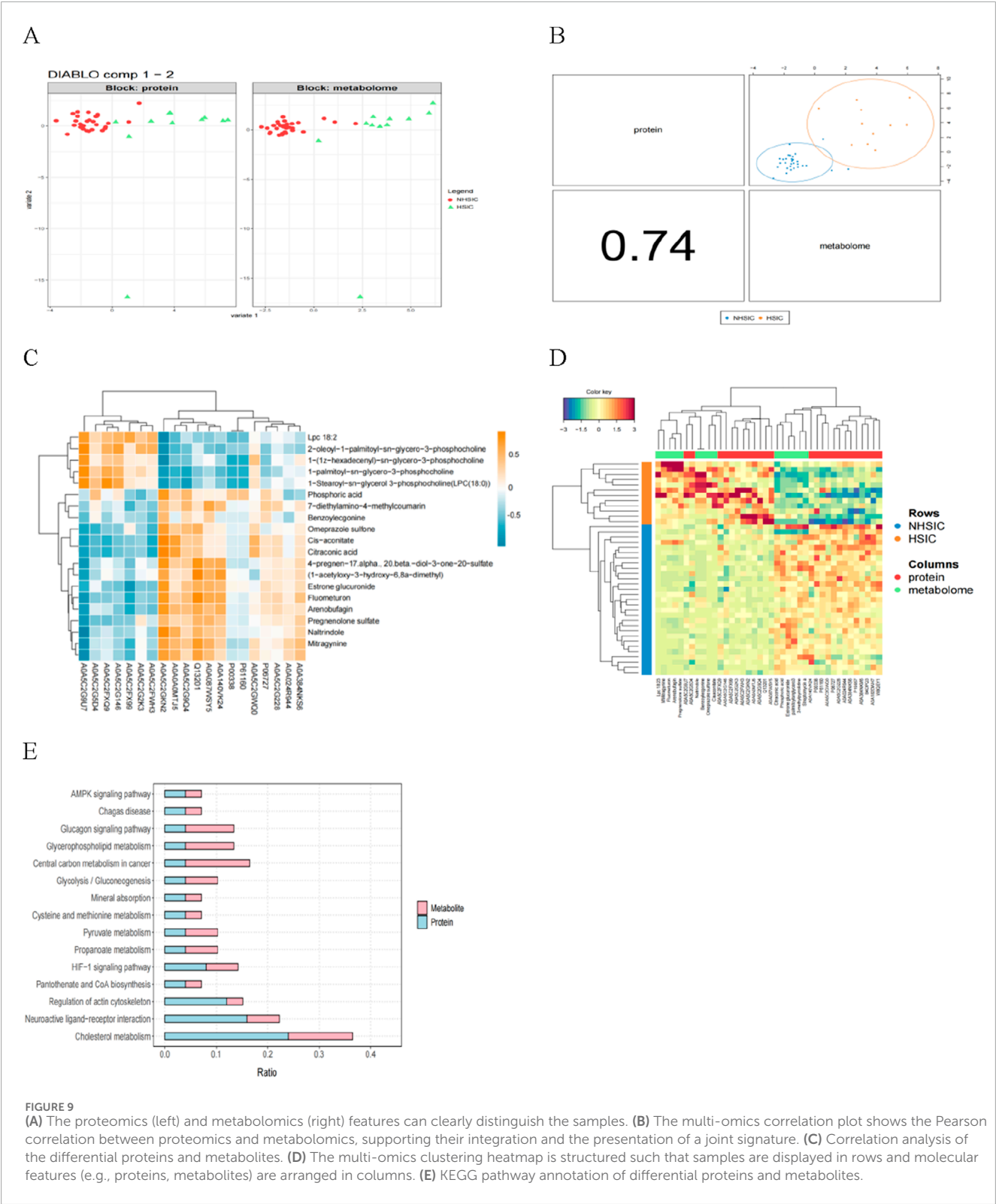


FIGURE 8 (A) Score chart of PCA analysis. (B) Score chart of OPLS-DA analysis. (C) PLS-DA model validation diagram. (D) A volcano plot of the differential metabolites. (E) A bubble diagram of top-25 metabolic pathways.

of these proteins emphasized the role of key signaling pathways, including complement and coagulation cascades. Heatstroke, a septic-like condition, involves these pathways, which are crucial in both sepsis and HSIC (Jiang et al., 2024; Liu et al., 2020b). Subsequently, a metabolomic analysis was performed and 103 DEMs were identified. Consistent with the proteomics results,



we observed significant regulation of energy metabolism-related biological pathways. Multiomics analysis of metabolomics and proteomics datasets based on the same biological samples was applied in this study. It was found that HS is closely associated with energy regulation-related pathways as well as lipid and carbohydrate

metabolism. Machine learning models showed better performance in diagnose HSIC, further supporting previous research in this area (Umemura et al., 2024; Zeng et al., 2023).

The main academic contribution of this study lies in providing new insights into the diagnosis of HSIC through the analysis

of clinical data and proteomics data, combined with machine learning methods. First, our study identified and analyzed 125 DEPs and 100DEMs, revealing significant differences in protein and metabolite expression between NHSIC and HSIC, particularly in pathways related to lipid and carbohydrate metabolism. Secondly, through LASSO regression and random forest analysis of these differentially expressed proteins, four key proteins -LDHA, NGAL, GBE, and Prothrombin -were identified. LASSO is a widely used regression method for high-dimensional data analysis that automatically selects variables through L1 regularization. It is particularly suitable when the number of features far exceeds the sample size, improving model predictive performance and enhancing interpretability. Boruta, a random forest-based feature selection algorithm, systematically evaluates feature importance to ensure optimal predictivity and interpretability. Combining LASSO with Boruta results in a more stable model with stronger interpretability. LDHA activation under hypoxia/inflammation drives lactate accumulation, which signals through endothelial GPR81 to increase vascular permeability and amplify systemic inflammation (Huau et al., 2024; Chen et al., 2023; Khatib-Massalha et al., 2020). Additionally, the metabolic product of LDHA, lactate, can affect endothelial cell permeability and coagulation function (Yang et al., 2022). NGAL, an acute-phase protein, is upregulated in inflammatory responses (Chakraborty et al., 2012). NGAL is not only involved in the activation of neutrophils but also promotes the vicious cycle of thrombus formation and inflammatory responses through its interaction with the coagulation system (Xue et al., 2025). Heatstroke can also cause renal damage, so its levels are correlated with the severity of heatstroke (Chen et al., 2024). GBE is closely related to lipid metabolism, and abnormalities in lipid metabolism during heatstroke may lead to changes in its levels. Prothrombin is a pivotal protein within the coagulation system. Its accelerated activation under inflammatory conditions can precipitate coagulation dysfunction, elevate the risk of thrombosis, and further intensify inflammatory responses (Teodoro et al., 2023). Moreover, growing evidence suggests that heparanase may contribute to heatstroke-induced coagulopathy by causing endothelial damage and glycocalyx degradation, which enhances tissue factor (TF) exposure, promotes prothrombin activation, and leads to excessive coagulation activation (Capozzi et al., 2021). Since heatstroke is frequently associated with coagulation abnormalities, changes in prothrombin are associated with HSIC (Degen and Sun, 1998; Wright et al., 1946). Based on these four important proteins, several machine learning models (XGBoost, LR, and SVM) were developed for the diagnosis of HIC. Comprehensive performance evaluations showed that the logistic regression (LR) model demonstrated the best diagnostic ability in clinical applications, with high sensitivity, specificity, and clinical utility, enabling early diagnosis and risk assessment of HSIC. However, this study has limitations that should be addressed in future research. The sample size of 41 patients is relatively small, which may limit the generalizability of the findings. Additionally, the study focused on a limited set of clinical and proteomic markers; further studies could explore a broader range of biomarkers and clinical variables. Longitudinal data is needed to validate the prognostic value of the identified biomarkers and the performance of the machine learning models in predicting long-term outcomes. Lastly, the interpretability of machine learning

models, while explored in this study, could benefit from further refinement to enhance clinical utility, ensuring that these models are both accurate and explainable for healthcare providers.

In summary, this study provides valuable insights into the clinical and molecular mechanisms of HSIC and demonstrates the utility of machine learning models in improving the diagnosis and diagnosis of HSIC. Future research should aim to validate these findings in larger cohorts and explore additional biomarkers that could further enhance the understanding and management of HSIC.

Data availability statement

The data presented in the study are deposited in the ProteomeXchange repository, accession number PXD064588.

Ethics statement

The studies involving humans were approved by the 908th Hospital of Chinese PLA Logistic Support Force. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

Author contributions

QZ: Data curation, Writing – original draft. QL: Formal Analysis, Writing – original draft. LH: Formal Analysis, Writing – original draft. LZ: Methodology, Writing – original draft. YZ: Writing – original draft, Methodology. XD: Visualization, Writing – original draft. NZ: Writing – original draft, Visualization. QS: Writing – review and editing, Supervision. JS: Supervision, Conceptualization, Writing – review and editing, Project administration.

Funding

The author(s) declare that financial support was received for the research and/or publication of this article. The study was funded by Chinese Medicine Education Association (No. 2022KTZ013). The funders were not involved in research design, data collection and manuscript preparation.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

References

- Author Anonymous (2021). Update: Heat illness, active component, U.S. Armed Forces, 2020. *MSMR* 28 (4), 10–15.
- Bouchama, A., Abuyassin, B., Lehe, C., Laitano, O., Jay, O., O'Connor, F. G., et al. (2022). Classic and exertional heatstroke. *Nat. Rev. Dis. Prim.* 8 (1), 8. doi:10.1038/s41572-021-00334-6
- Bouchama, A., and Knochel, J. P. (2002). Heat stroke. *N. Engl. J. Med.* 346, 1978–1988. doi:10.1056/NEJMra011089
- Capozzi, A., Riitano, G., Recalchi, S., Manganello, V., Costi, R., Saccoliti, F., et al. (2021). Effect of heparanase inhibitor on tissue factor overexpression in platelets and endothelial cells induced by anti- β 2-GPI antibodies. *J. Thromb. Haemost.* 19 (9), 2302–2313. doi:10.1111/jth.15417
- Caputa, M., Dokladny, K., and Kurowicka, B. (2000). Endotoxaemia does not limit heat tolerance in rats: the role of plasma lipoproteins. *Eur. J. Appl. Physiol.* 82 (1–2), 142–150. doi:10.1007/s004210050664
- Chakraborty, S., Kaur, S., Guha, S., and Batra, S. K. (2012). The multifaceted roles of neutrophil gelatinase associated lipocalin (NGAL) in inflammation and cancer. *Biochim. Biophys. Acta* 1826 (1), 129–169. doi:10.1016/j.bbcan.2012.03.008
- Chen, L., Liu, C., Zhang, Z., Zhang, Y., and Feng, X. (2024). Effects of normal saline versus lactated Ringer's solution on organ function and inflammatory responses to heatstroke in rats. *J. Intensive Care* 12 (1), 39. doi:10.1186/s40560-024-00746-y
- Chen, Y., Pan, Z., Bai, Y., and Xu, S. (2023). Redox state and metabolic responses to severe heat stress in lenok *Brachymystax lenok* (Salmonidae). *Front. Mol. Biosci.* 10, 1156310. doi:10.3389/fmolb.2023.1156310
- Degen, S. J., and Sun, W. Y. (1998). The biology of prothrombin. *Crit. Rev. Eukaryot. Gene Expr.* 8 (2), 203–224. doi:10.1615/critrevueukargeneexpr.v8.i2.60
- He, L., Lin, Q., Zhong, L., Zeng, Q., and Song, J. (2022). Thromboelastography maximum amplitude as an early predictor of disseminated intravascular coagulation in patients with heatstroke. *Int. J. Hyperth.* 39 (1), 605–610. doi:10.1080/02656736.2022.2066206
- Huau, G., Liaubet, L., Gourdiere, J. L., Riquet, J., and Renaudeau, D. (2024). Multi-tissue metabolic and transcriptomic responses to a short-term heat stress in swine. *BMC Genomics* 25 (1), 99. doi:10.1186/s12864-024-09999-1
- Huise, M. G., Pease, S., Hurtado-Nedelec, M., Arnaud, B., Malaquin, C., Wolff, M., et al. (2008). Leukocyte activation: the link between inflammation and coagulation during heatstroke. A study of patients during the 2003 heat wave in Paris. *Crit. Care Med.* 36, 2288–2295. doi:10.1097/CCM.0b013e318180dd43
- Iba, T., Connors, J. M., Levi, M., and Levy, J. H. (2022). Heatstroke-induced coagulopathy: biomarkers, mechanistic insights, and patient management. *EClinicalMedicine* 44, 101276. doi:10.1016/j.eclinm.2022.101276
- Iba, T., Kondo, Y., Maier, C. L., Helms, J., Ferrer, R., and Levy, J. H. (2025). Impact of hyper- and hypothermia on cellular and whole-body physiology. *J. Intensive Care* 13 (1), 4. doi:10.1186/s40560-024-00774-8
- Jiang, H., Guo, Y., Wang, Q., Wang, Y., Peng, D., Fang, Y., et al. (2024). The dysfunction of complement and coagulation in diseases: the implications for the therapeutic interventions. *MedComm* 5 (11), e785. doi:10.1002/mco2.785
- Khatib-Massalha, E., Bhattacharya, S., Massalha, H., Biram, A., Golan, K., Kollet, O., et al. (2020). Lactate released by inflammatory bone marrow neutrophils induces their mobilization via endothelial GPR81 signaling. *Nat. Commun.* 11 (1), 3547. doi:10.1038/s41467-020-17402-2
- Li, Y., Li, H., Ma, W., Maegele, M., Tang, Y., and Gu, Z. (2023). Proteomic profiling of serum exosomes reveals acute phase response and promotion of inflammatory and platelet activation pathways in patients with heat stroke. *PeerJ* 11, e16590. doi:10.7717/peerj.16590
- Lin, Q. W., Zhong, L. C., He, L. P., Zeng, Q. B., Zhang, W., Song, Q., et al. (2024). A newly proposed heatstroke-induced coagulopathy score in patients with heat illness: a multicenter retrospective study in China. *Chin. J. Traumatol.* 27 (2), 83–90. doi:10.1016/j.cjte.2023.08.001
- Liu, J., Li, Q., Zou, Z., Li, L., and Gu, Z. (2025). The pathogenesis and management of heatstroke and heatstroke-induced lung injury. *Burns Trauma* 13, tkae048. doi:10.1093/burnst/tkae048
- Liu, S. Y., Song, J. C., Mao, H. D., Zhao, J. B., Song, Q., and Expert Group of Heat Stroke Prevention and Treatment of the People's Liberation Army (2020a). Expert consensus on the diagnosis and treatment of heat stroke in China. *Mil. Med. Res.* 7 (1), 1. doi:10.1186/s40779-019-0229-2
- Liu, S. Y., Wang, Q., Lou, Y. P., Gao, Y., Ning, B., Song, Q., et al. (2020b). Interpretations and comments for expert consensus on the diagnosis and treatment of heat stroke in China. *Mil. Med. Res.* 7 (1), 37. doi:10.1186/s40779-020-00266-4
- Ogata, S., Takegami, M., Ozaki, T., Nakashima, T., Onozuka, D., Murata, S., et al. (2021). Heatstroke predictions by machine learning, weather information, and an all-population registry for 12-hour heatstroke alerts. *Nat. Commun.* 12, 4575. doi:10.1038/s41467-021-24823-0
- Shimazaki, J., Hifumi, T., Shimizu, K., Oda, Y., Kanda, J., Kondo, Y., et al. (2020). Clinical characteristics, prognostic factors, and outcomes of heat-related illness (Heatstroke Study 2017–2018). *Acute Med. Surg.* 7 (1), e516. doi:10.1002/ams2.516
- Teodoro, A. G. F., Rodrigues, W. F., Farnesi-de-Assunção, T. S., Borges, A. V. B. E., Obata, M. M. S., Neto, J. R. C., et al. (2023). Inflammatory response and activation of coagulation after COVID-19 infection. *Viruses* 15 (4), 938. doi:10.3390/v15040938
- Umehura, Y., Okada, N., Ogura, H., Oda, J., and Fujimi, S. (2024). A machine learning model for early and accurate prediction of overt disseminated intravascular coagulation before its progression to an overt stage. *Res. Pract. Thromb. Haemost.* 8 (5), 102519. doi:10.1016/j.rpth.2024.102519
- Wright, D. O., Reppert, L. B., and Cuttino, J. T. (1946). Purpuric manifestations of heatstroke; studies of prothrombin and platelets in 12 cases. *Arch. Intern. Med. (Chic)* 77, 27–36. doi:10.1001/archinte.1946.00210360032003
- Xue, M., Wang, S., Li, C., Wang, Y., Liu, M., Huang, X., et al. (2025). Deficiency of neutrophil gelatinase-associated lipocalin elicits a hemophilia-like bleeding and clotting disorder in mice. *Blood* 145 (9), 975–987. doi:10.1182/blood.2024026476
- Yang, K., Fan, M., Wang, X., Xu, J., Wang, Y., Gill, P. S., et al. (2022). Lactate induces vascular permeability via disruption of VE-cadherin in endothelial cells during sepsis. *Sci. Adv.* 8 (17), eabm8965. doi:10.1126/sciadv.abm8965
- Zeng, Q., Zhong, L., Zhang, N., He, L., Lin, Q., and Song, J. (2023). Nomogram for predicting disseminated intravascular coagulation in heatstroke patients: a 10 years retrospective study. *Front. Med. (Lausanne)* 10, 1150623. doi:10.3389/fmed.2023.1150623
- Zhong, L., Wu, M., Wang, C., Yu, B., Liu, Z., and Liu, Z. (2021). Clinical characteristics and outcomes of patients with severe heatstroke complicated with disseminated intravascular coagulation: a case-control study. *Thromb. Res.* 197, 120–123. doi:10.1016/j.thromres.2020.11.009