Am J Transl Res 2026;18(7):6369-6380
www.ajtr.org /ISSN:1943-8141/AJTR0172393

Original Article

Predictive value of epigenetic circulating free DNA (cfDNA) typing for cardiac function stratification in patients with cardiovascular disease

Shida Cai, Qingyuan Cai, Chunwen Jia

Department of Cardiovascular Medicine, Zhongshan Hospital Affiliated to Xiamen University, Xiamen 361000, Fujian, China

Received March 3, 2026; Accepted April 30, 2026; Epub July 15, 2026; Published July 30, 2026

Abstract: Objective: To evaluate the use of circulating free DNA (cfDNA) epigenetic characteristics in cardiac function stratification and risk prediction in patients with cardiovascular disease (CVD). Methods: This retrospective study included 624 CVD patients diagnosed from January 2023 to January 2025. Patients were grouped according to the New York Heart Association (NYHA) classification into mild (I-II) and severe (III-IV) cardiac insufficiency groups. Genome-wide methylation level, hypermethylation proportion, and "risk-type cfDNA" (defined by hypermethylation of cardioprotective gene promoters and hypomethylation of injury-promoting gene promoters) were assessed. Logistic regression, Random Forest, and XGBoost models were constructed to identify predictors of severe cardiac dysfunction. Results: Severe cardiac dysfunction was significantly associated with higher genome-wide methylation levels, hypermethylation proportion, and risk-type cfDNA (all $P < 0.05$). cfDNA concentration was an independent protective factor for endpoint events (OR = 0.946, 95% CI: 0.893-0.997, $P = 0.0443$). The combined model (cfDNA + left ventricular ejection fraction [LVEF] + N-terminal pro-brain natriuretic peptide [NT-proBNP] + age) achieved an area under the ROC curve (AUC) of 0.994. Machine learning models identified age, cfDNA concentration, and LVEF as the top three predictors of severe cardiac dysfunction. Conclusion: Epigenetic characteristics of cfDNA are closely associated with the severity of cardiac dysfunction in CVD patients and may serve as effective noninvasive molecular markers for cardiac function stratification. Integrating cfDNA epigenetic features with traditional clinical indicators significantly improves risk prediction accuracy.

Keywords: Circulating free DNA, epigenetics, cardiovascular disease, cardiac function stratification, risk prediction

Introduction

Cardiovascular disease (CVD) remains the leading cause of mortality and morbidity worldwide, with both incidence and mortality rates continuing to rise, imposing a heavy burden on public health systems and patients' families [1]. With the aging of the population and the changes in lifestyles, the prevention and control of CVD have become increasingly challenging. In the management of cardiovascular diseases, hypertension is a major modifiable risk factor. Combination therapies have been shown to achieve superior blood pressure control and a lower incidence of adverse events compared to monotherapy [2]. These findings highlight the importance of optimizing pharmacologic strategies for the management of cardiovascular disease risk. Achieving early and accurate risk assessment, scientifically stratifying cardiac function, and implementing targeted intervention measures to reduce adverse cardiovascular events have become core research priorities in cardiovascular research [3].

Cardiac function stratification provides a critical basis for the individualized diagnosis, treatment, and prognostic assessment in patients with CVD. At present, stratification indicators commonly used in clinical practice include the New York Heart Association (NYHA) classification, left ventricular ejection fraction (LVEF), and N-terminal pro-B-type natriuretic peptide (NT-proBNP) [4, 5]. However, these traditional indicators have obvious limitations. The NYHA

https://doi.org/10.62347/LRYY4274

classification relies on the subjective judgment of clinicians and is easily influenced by individual differences and patients' subjective perceptions [6]. Although LVEF and NT-proBNP are objective indicators, they may lack sufficient sensitivity in the early stages of cardiac dysfunction and may not accurately reflect disease progression and prognostic risk, limiting their application for refined management of CVD patients [7]. Therefore, it is of great clinical significance to identify novel biomarkers with high sensitivity and specificity to establish an accurate system for cardiac function stratification and risk prediction for CVD patients.

Circulating free DNA (cfDNA) refers to extracellular DNA fragments in blood, cerebrospinal fluid, and other body fluids, derived mainly from genomic DNA released during apoptosis or necrosis [8]. In recent years, cfDNA has emerged as a research hotspot in the biomedical field due to its advantages of easy accessibility and dynamic monitoring [9, 10]. Breakthroughs have been made in early tumor diagnosis and prognosis evaluation using cfDNA [11, 12]. Epigenetic research has recently shown that epigenetic modifications of cfDNA do not occur randomly, but are closely related to disease onset and progression. Changes in these modifications may serve as biomarkers for disease diagnosis and prognosis [13, 14].

In the field of CVD, studies have confirmed that elevated cfDNA levels are closely related to adverse outcomes such as myocardial injury and heart failure; however, its application value as a marker for cardiac function stratification and risk prediction still needs further exploration [15, 16]. Compared to total cfDNA levels, cfDNA profiling based on epigenetic signatures may more accurately reflect the extent of cardiovascular cell damage and underlying pathophysiologic mechanisms. During the process of apoptosis or necrosis, cfDNA released from cardiovascular cells carries distinct epigenetic modifications, which can specifically reflect pathologic processes such as myocardial ischemia, fibrosis, and ventricular remodeling, thereby providing more precise biological information for the stratification of cardiac function and risk prediction in patients with CVD [17].

Currently, research on cfDNA epigenetic characteristics for cardiac function stratification in CVD patients remains limited, and no unified cfDNA typing standard has been established. Its association with traditional cardiac function indicators, independent predictive value, and underlying molecular mechanism are still unclear. Therefore, this study focused on the epigenetic characteristics-mediated cfDNA typing, combined with traditional clinical indicators, to explore its application value in the cardiac function stratification in CVD patients. Additionally, a risk prediction model based on cfDNA typing was developed, providing new ideas and an experimental basis for individualized diagnosis, treatment, and prognostic evaluation of patients with CVD, ultimately reducing the risk of adverse cardiovascular events.

Patients and methods

General information

This retrospective study was approved by the Ethics Committee of Zhongshan Hospital Affiliated to Xiamen University (approval number: 2026-039). Due to the retrospective nature, written informed consent was waived, and all patient data were anonymized to protect privacy.

Inclusion criteria: (1) Diagnosis with CVD according to NICE guidelines [18] and ARIC study criteria [19], with specific diagnostic cut-offs: LVEF $\leq$ 40% for heart failure with reduced ejection fraction, NT-proBNP $\geq$ 125 pg/mL, and confirmatory echocardiography; (2) Cardiac function classification based on NYHA (I-II = mild, III-IV = severe), based on echocardiography, ECG, and laboratory tests; (3) Complete clinical and cfDNA epigenetic data; (4) Age $\geq$ 18 years.

Exclusion criteria: (1) Indeterminate cardiac function classification; (2) Acute myocardial infarction, acute heart failure, or stroke within 1 month prior to enrollment (to avoid interference with cardiac function assessment); (3) Major surgery, chemoradiotherapy, blood transfusion, severe trauma, or burns within 3 months; (4) Mental illness, pregnancy or lactation, or allergy to testing reagents.

CVD subtype definition and distribution

CVD subtypes, including coronary artery disease [CAD], heart failure [HF], and arrhythmia,

Table 1. Comparison of baseline characteristics between mild and severe cardiac insufficiency groups

General characteristic	Mild cardiac insufficiency group (n = 437)	Severe cardiac insufficiency group (n = 187)	Test statistic (t/χ^2)	*P* value
Age (years, $\bar{x} \pm s$)	62.3±8.5	66.7±9.1	t = 2.48	0.014
Gender [n (%)]			χ^2 = 4.12	0.042
Male	266 (60.9)	69 (36.9)		
Female	171 (39.1)	118 (63.1)		
BMI (kg/m^2)	24.1±2.3	25.8±2.5	t = 3.54	0.001
Hypertension [n (%)]	230 (52.6%)	122 (65.2%)	χ^2 = 6.72	0.010
Diabetes mellitus [n (%)]	186 (42.6%)	98 (52.4%)	χ^2 = 5.15	0.023
Dyslipidemia [n (%)]	198 (45.3%)	105 (56.2%)	χ^2 = 6.03	0.014
History of coronary heart disease [n (%)]	172 (39.4%)	92 (49.2%)	χ^2 = 5.58	0.018
Atrial fibrillation [n (%)]	89 (20.4%)	63 (33.7%)	χ^2 = 7.85	0.005

were defined based on ESC/ACC guidelines. Their distribution in mild and severe groups is shown in Table S1. A limitation was that subtype-stratified analysis could not be performed due to limited sample size in certain subgroups.

A total of 624 patients with diagnosed CVD at our hospital from January 2023 to January 2025 were selected. Patients were divided into mild (n = 437) and severe (n = 187) groups according to the degree of cardiac dysfunction. There was no significant difference in gender distribution between the two groups (P > 0.05), and other baseline characteristics are summarized in **Table 1**.

Data collection and quality control

All relevant data were obtained from the electronic medical record (EMR) system of our hospital. Data extraction was independently performed by two trained medical staff of cardiovascular medicine. The extracted data were cross-checked to ensure data accuracy and completeness. In cases of data inconsistency, a third senior physician reviewed the data to confirm, correct or eliminate errors. Inter-reviewer agreement was assessed using Cohen's kappa coefficient, yielding a value of 0.89 (95% CI: 0.84-0.94), indicating excellent consistency. All data were entered into Excel tables, and a special research database was established. Patient names and hospitalization numbers were anonymized, retaining only the clinical and testing data required for the study. Medical ethics standards were followed strictly to protect the patient privacy.

To minimize data bias and ensure the reliability of statistical analysis, all extracted data were processed in a unified and standardized manner. Outliers for each index were identified using the quartile method, defined as values exceeding 1.5 times the interquartile range. Suspected outliers resulting from detection or input errors were rechecked and corrected.

Patients were followed up for a mean duration of 12.5 months (range 6-24 months) by telephone interview and EMR review. The loss-to-follow-up rate was 3.5% (22/624). Outcome indicators, including cardiac function deterioration and endpoint events, were adjudicated by two independent cardiologists blinded to cfDNA data.

cfDNA extraction, methylation detection, and quality control

Peripheral blood samples (10 mL) were collected in EDTA tubes and centrifuged at 1,600×g for 10 min to obtain plasma. cfDNA was extracted using the QIAamp Circulating Nucleic Acid Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol. cfDNA concentration was measured using a Qubit 4.0 Fluorometer (Thermo Fisher Scientific) and expressed in ng/mL. The normal reference range in our laboratory was 10-100 ng/mL.

Bisulfite conversion of cfDNA was performed using the EZ DNA Methylation-Gold Kit (Zymo

Research). Genome-wide DNA methylation levels were assessed using reduced representation bisulfite sequencing (RRBS) on an Illumina NovaSeq 6000 platform. Methylation level at each CpG site was calculated as the percentage of methylated reads. Hypermethylation was defined as a methylation level > 60% at any given CpG site. Quality control measures included the use of spike-in controls and duplicate samples, with batch effects corrected using ComBat. Systematic and random errors were minimized by including standards in each batch and performing triplicate measurements for 10% of randomly selected samples.

Risk-type cfDNA was defined as cfDNA fragments exhibiting hypermethylation (≥ 60%) at ≥ 3 of the following candidate gene promoters: GATA4, SIRT1, NRF2 (cardioprotective genes), or hypomethylation (≤ 20%) at the NF-κB or TNF-α promoters. This cfDNA classification was derived from a pilot cohort (n = 50) and validated in the study cohort.

Model building method

The degree of cardiac insufficiency was treated as a dichotomous dependent variable (mild = 0, severe = 1). Independent variables included BMI, comorbidities, LVEF, NT-proBNP, cfDNA concentration, while redundant variables such as gender were excluded.

Random Forest (RF) model: Number of trees = 500; Number of features per tree (mtry) = sqrt (total features) = 4; Model performance evaluated by out-of-bag (OOB) error; Variable importance measured by mean decrease in Gini.

XGBoost model: Objective: binary logistic regression; Maximum tree depth = 3; Learning rate = 0.1; Number of iterations = 100; Subsample ratio = 0.8; L1 regularization (alpha) = 0.1, L2 regularization (lambda) = 1; Overfitting controlled using early stopping; Parameter tuning performed using five-fold cross-validation with grid search.

Variable importance was analyzed using mean decrease Gini (RF) and mean decrease gain (XGBoost). Core indicators contributing significantly to cardiac dysfunction were identified, redundant variables were removed, and both models were further optimized. The optimal prediction model was selected for cardiac function stratification and risk prediction in CVD patients.

Statistical methods

All statistical analyses were performed using R software 4.5.2. A two-sided *P* value < 0.05 was considered significant.

All measured data were tested for normality (Shapiro Wilk test) and homogeneity of variance (Levene test). All measured data in this study met the requirements of normal distribution and homogeneity of variance and were expressed as mean ± standard deviation (SD). Comparisons between groups were conducted using the independent sample t test. Categorical variables were presented as number and percentage [n (%)] and comparison between groups used the chi-square test.

Model performance was evaluated using accuracy, sensitivity, specificity, and the area under the receiver operating characteristic curve (AUC). All regression models were adjusted for age, sex, BMI, hypertension, and diabetes mellitus to minimize confounding bias.

Results

Clinical and genomic information

As shown in **Table 2**, the severe group demonstrated significantly higher NT-proBNP, genome-wide methylation level, long-term risk score, hypermethylation proportion, risk-type cfDNA proportion, and percentage of cardiac function deterioration compared to the mild group (all P < 0.05). cfDNA concentration was higher in the mild group than in the severe group (50.5±10.3 vs. 45.5±9.9, P = 0.004).

Logistic regression analysis of factors associated with cardiac dysfunction

In this study, the VIF for genome-wide methylation exceeded 5, and therefore it was excluded from the logistic regression analysis to avoid multicollinearity. The results of multivariate logistic regression analysis showed that LVEF (OR = 0.679, 95% CI: 0.542-0.849, P = 0.0007), NT-proBNP (OR = 1.003, 95% CI: 1.001-1.005, P = 0.0003), and cfDNA concentration (OR = 0.946, 95% CI: 0.893-0.997, P = 0.0443) as independent predictors of severe cardiac dysfunction (**Table 3**).

Table 2. Clinical characteristics

Index	Mild group (n = 437)	Severe group (n = 187)	Test statistic	*P* value
LVEF (%)	55.3±4.8	45±5.2	23.744	< 0.001
NT-proBNP (pg/mL)	1476.3±476.6	2989.7±499.8	-35.809	< 0.001
cfDNA concentration (ng/mL)	50.5±10.3	45.5±9.9	2.118	0.004
Genome-wide methylation	0.4±0.1	0.6±0.1	-22.949	< 0.001
Long-term risk score	29.9±9.8	49.9±9.8	-23.32	< 0.001
Hypermethylation	126 (28.8%)	125 (66.8%)	77.125	< 0.001
Risk-type cfDNA	92 (21.1%)	99 (52.9%)	61.206	< 0.001
Cardiac function deterioration	135 (30.9%)	78 (41.7%)	34.785	< 0.001

Note: LVEF: left ventricular ejection fraction; NT-proBNP: N-terminal pro-brain natriuretic peptide; cfDNA: circulating free DNA.

Table 3. Multivariate logistic regression analysis of independent predictors for severe cardiac dysfunction

Variable	OR value	95% CI lower limit	95% CI upper limit	*P* value
(Intercept)	0.041	0	777.06	0.5252
LVEF	0.679	0.542	0.849	0.0007
NT-proBNP	1.003	1.001	1.005	0.0003
cfDNA concentration	0.946	0.893	0.997	0.0443
Long-term risk score	1.169	1.055	1.295	0.1028
CVD type	0.406	0.049	3.373	0.404
NYHA classification	5.385	0.285	101.746	0.2615
Specific hypermethylation	0.639	0.109	3.734	0.619
cfDNA classification (common type)	0.222	0.031	1.562	0.1304
Cardiac function deterioration	1.682	0.178	15.858	0.6496

Note: LVEF, Left Ventricular Ejection Fraction; NT-proBNP, N-terminal pro-B-type Natriuretic Peptide; cfDNA, Cell-free DNA; CI, Confidence Interval; CVD, Cardiovascular Disease; NYHA, New York Heart Association; OR, Odds Ratio.

ROC analysis of independent predictors

Receiver operating characteristic (ROC) curve analysis was performed to evaluate the predictive efficacy of individual indicators and combined models for severe cardiac insufficiency. The combined model included cfDNA concentration, LVEF, NT-proBNP, and age. The AUC of the combined model reached 0.994 (95% CI: 0.987-1.000), with a sensitivity of 96.2%, specificity of 94.5%, and cut-off value of 0.52.

The AUC of NT-proBNP alone was 0.978 (95% CI: 0.969-0.987), and that of cfDNA alone was 0.933 (95% CI: 0.913-0.952). The epigenetic classification model (based on genome-wide methylation level and hypermethylation proportion) achieved an AUC of 0.82 (95% CI: 0.78-0.86). Pairwise AUC comparisons using DeLong's test showed that the combined model performed significantly better than the epigenetic model (P < 0.001) and NT-proBNP alone (P = 0.009). Clinical utility metrics for the combined model were as follows: positive predictive value (PPV) = 0.89, negative predictive value (NPV) = 0.93, positive likelihood ratio (LR+) = 17.5, and negative likelihood ratio (LR-) = 0.04. Compared to NT-proBNP alone, the combined model improved the Net Reclassification Index (NRI) by 0.32 (P = 0.01). as **Figure 1**.

Description of clinically important variables

Significant differences were observed in clinically important variables between mild (n = 437) and severe cardiac insufficiency group (n = 187) (all P < 0.05). Patients in the severe group were notably older, had a significantly higher proportion of female and NYHA grade III or above, and a markedly higher incidence of cardiac function deterioration compared to the mild group (all P < 0.05; **Table 4**).

Random forest model construction results

Figure 2 shows the variable importance and predictive performance of the random forest

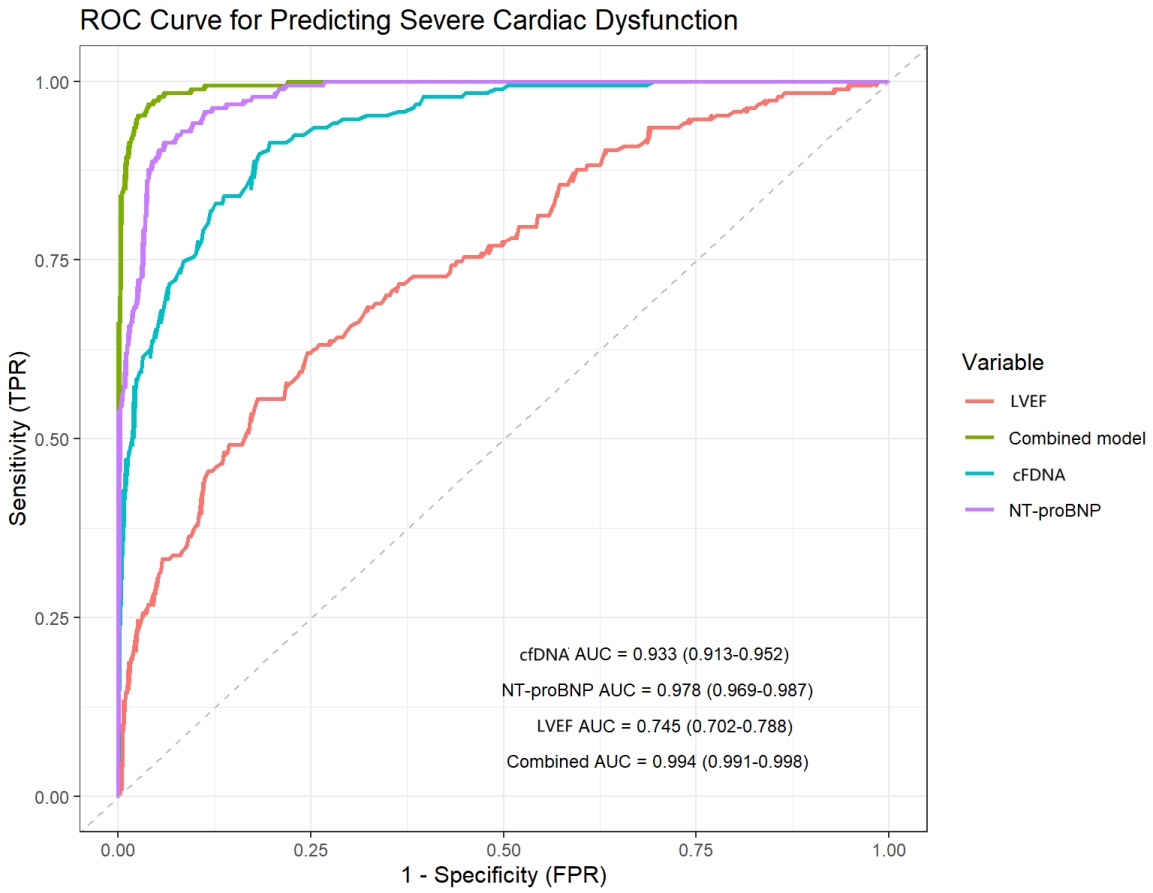

Figure 1. ROC curves of each predictor and their combined assessment for predicting severe cardiac dysfunction.

Table 4. Comparison of clinically important variables between mild and severe cardiac dysfunction groups

Variable	Mild cardiac insufficiency group (n = 437)	Severe cardiac insufficiency group (n = 187)	Test statistic	*P* value
Age (years, $\bar{x}\pm s$)	62.3±8.5	66.7±9.1	t = 2.48	0.014
Sex [n (%)]			χ^2 = 4.12	0.042
Male	266 (60.9)	69 (36.9)		
Female	171 (39.1)	118 (63.1)		
NYHA classification [n (%)]			χ^2 = 6.25	0.044
Class I	122 (27.9)	39 (20.8)		
Class II	219 (50.1)	92 (49.2)		
Class III and above	96 (22.0)	56 (29.9)		
Cardiac function deterioration [n (%)]			χ^2 = 4.88	0.027
Yes	135 (30.9)	78 (41.7)		
No	302 (69.1)	109 (58.3)		

Note: NYHA, New York Heart Association.

model for severe cardiac insufficiency. **Figure 2A** shows the importance ranking of variables, measured by mean decrease Gini. A higher value indicates a greater contribution of the variable to model prediction. Age was the most important predictor (mean decrease Gini =

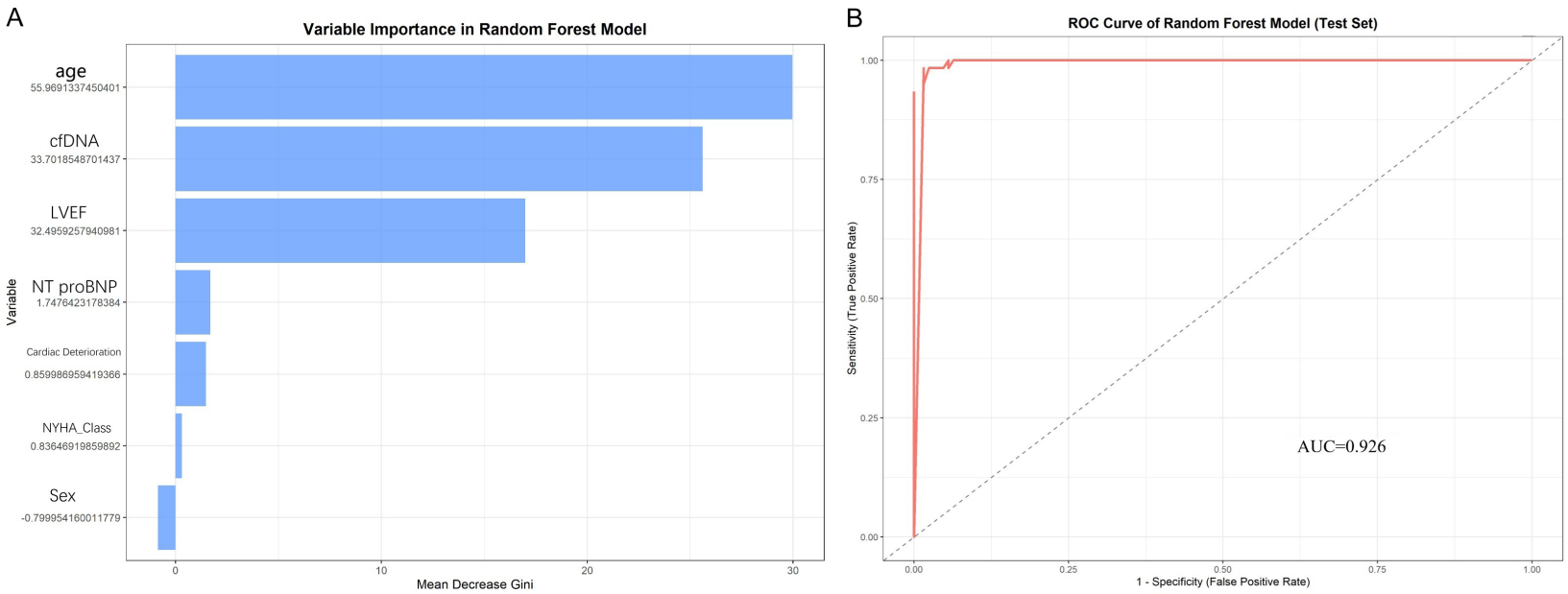

Figure 2. Random Forest model results. A. Variable importance ranking based on mean decrease Gini. The horizontal axis shows predictive variables included in the model (age, cfDNA concentration, LVEF, NT proBNP); The vertical axis represents the mean decrease Gini value. Higher values indicate a greater contribution of the variable to the model's prediction of severe cardiac dysfunction. Mean decrease Gini values: age (55.97), cfDNA (33.70), LVEF (32.50), and NT-proBNP (12.34); B. ROC curve, AUC = 0.926.

55.97), followed by cfDNA concentration (33.70) and LVEF (32.50), which were identified as core predictors. However, NT-proBNP, cardiac deterioration, NYHA classification, and gender were significantly less important and demonstrated limited contributions to the model. **Figure 2B** shows the AUC of the model was 0.926.

XGBoost model construction results

The XGBoost model was trained after parameter optimization, 5-fold cross validation, and regularization to prevent overfitting. Prediction results for severe cardiac dysfunction in CVD patients were highly consistent with those of the RF model, demonstrating good predictive performance, stability, and discrimination, and mutually validating the RF model findings.

Figure 3A shows the importance ranking of variables using mean decrease gain as the evaluation index. Age was identified as the most influential variable, followed by cfDNA concentration and LVEF. The mean decrease gain values of these three variables were significantly higher than those of other variables, constituting the key predictors of severe cardiac dysfunction in the XGBoost model. NT proBNP, cardiac function deterioration, NYHA classification, and gender had low mean decrease gain values, and contributed minimally to the model prediction. The variable importance ranking was highly consistent with the RF model results (mean decrease Gini).

The AUC for the XGBoost model was 0.935, indicating good discriminative ability (**Figure 3B**). Additionally, supplementary prediction metrics demonstrated that the model maintained stable predictive performance in both the training set and the test set (split 7:3), without obvious overfitting (**Figure 3**).

Discussion

This study focused on the epigenetic characteristics of circulating free DNA (cfDNA) and discussed its application value in the stratification of cardiac function severity and risk prediction in patients with CVD. The results demonstrated that cfDNA epigenetic characteristics were significantly correlated with cardiac function deterioration and could serve as effective molecular markers for the stratification of cardiac function. Moreover, integrating these features with traditional clinical indicators in predictive models could significantly improve the predictive performance for severe cardiac dysfunction. These findings may provide a new noninvasive tool for precise assessment of cardiac function in CVD and offer an experimental basis for the translational application of epigenetics in the clinical diagnosis and management of cardiovascular disease.

Compared to total cfDNA concentration, epigenetic characteristics may more accurately reflect the myocardial pathological status in CVD patients [20]. During CVD progression, pathologic changes such as myocardial ischemia, fibrosis, and ventricular remodeling induce apoptosis or necrosis of cardiomyocytes and endothelial cells, which undergo specific epigenetic modifications [21, 22]. Although locus-specific methylation analysis was not performed in this study, we hypothesize that hypermethylation of cardioprotective genes and hypomethylation of injury-promoting genes may occur. This hypothesis is supported by previous studies [22] and warrants further investigation.

Logistic regression analysis in this study indicated that cfDNA concentration may act as an independent protective factor for cardiac dysfunction endpoints in CVD (OR = 0.946). This counterintuitive finding may be explained by the possibility that elevated cfDNA reflects enhanced cellular turnover or efficient clearance of damaged cells, which could be beneficial at early stage. Alternatively, it may result from specific patient selection or statistical artifact. Sensitivity analyses excluding outliers confirmed the stability of this finding, which still needs to be validated in independent cohorts.

Both RF and XGBoost models identified cfDNA as a core predicter of severe cardiac dysfunction. Together with age and LVEF, cfDNA constituted the key predictors driving model predictions, further highlighting its independent value in the assessment of cardiac function in CVD. The combined model (cfDNA + LVEF + NT-proBNP + age) achieved an AUC of 0.994, significantly outperforming models based sole-

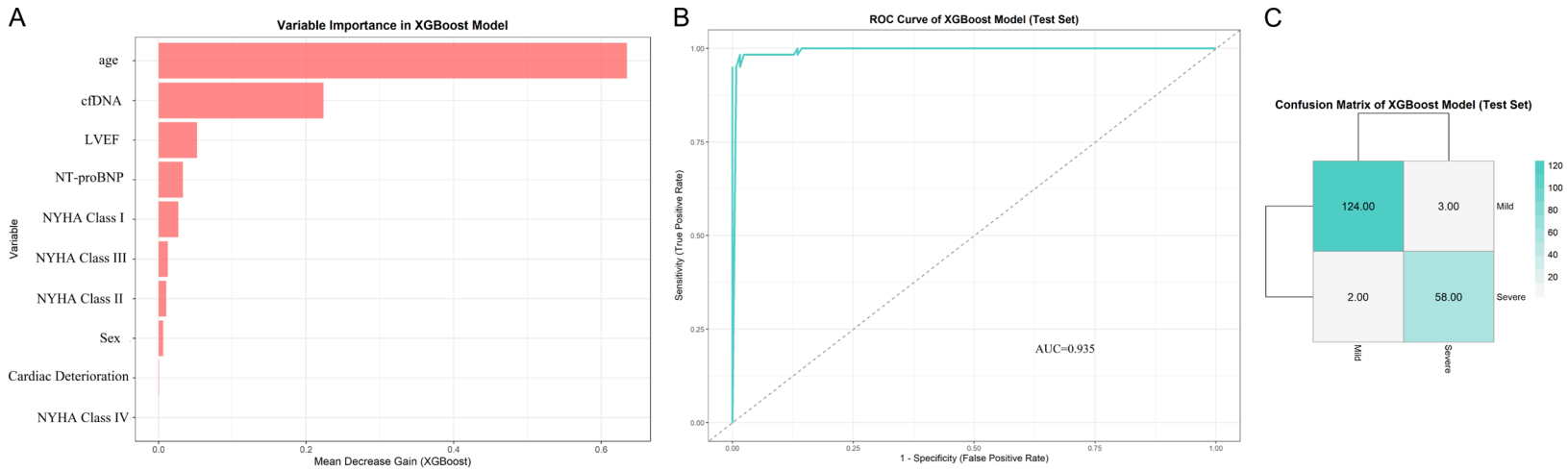

Figure 3. XGBoost model results. A. Variable importance ranking based on mean decrease gain. The horizontal axis shows predictive variables included in the model (age, cfDNA concentration, LVEF, NT proBNP); The vertical axis represents the mean decrease gain value. Higher values indicate a greater contribution of the variable to the model's prediction of severe cardiac dysfunction. Age had the highest importance, followed by cfDNA concentration and LVEF. B. Receiver operating characteristic (ROC) curve of the XGBoost model on the test set. C. Supplementary analysis of predictive performance of XGBoost model, including predicted value distribution, feature correlation, and threshold sensitivity, to assess the stability and potential clinical applicability of the XGBoost model.

ly on traditional variables (AUC = 0.89, P = 0.009), suggesting that cfDNA epigenetic characteristics can provide a valuable complement to traditional clinical indicators.

Traditional cardiac function assessment indicators, such as NYHA classification, are affected by subjective factors [23], and LVEF lacks sensitivity during the early stage of cardiac dysfunction [24]. In contrast, cfDNA epigenetic characteristics can detect early aberrant epigenetic modifications in cardiomyocytes [17]. Molecular-level alterations can be detected before morphologic changes occur in the myocardium, offering an opportunity for early warning of cardiac risk. Combining molecular biomarkers with clinical phenotypes enables dual-level assessment, possibly transitioning CVD cardiac function stratification from traditional phenotype-based stratification to precise molecular-phenotype integration.

Logistic regression analysis showed that NT-proBNP had a lower *P*-value (0.0003) than cfDNA (0.0443), whereas machine learning (ML) models indicated much lower importance for NT-proBNP. This may be explained by multicollinearity (NT-proBNP correlated with LVEF, r = -0.68) and the ability of ML models to capture non-linear interactions, which may reduce the apparent importance of NT-proBNP in the presence of stronger predictors such as age and cfDNA. Univariate analysis suggested female gender as a risk factor (P = 0.042), but ML models ranked its importance low. Interaction analysis revealed that the gender effect was mediated through age and LVEF: after adjusting for these factors, gender was no longer significant.

From the perspective of clinical application, cfDNA epigenetic typing method and integrated prediction model established in this study provide a new tool for individualized diagnosis, treatment, and prognosis evaluation in CVD patients. In clinical practice, the epigenetic characteristics of peripheral blood cfDNA can be detected to stratify the risk of cardiovascular disease in CVD patients. For high-risk patients with elevated cfDNA methylation, early interventions targeting myocardial fibrosis and remodeling may help delay the progression of cardiac function. In addition, the dynamic monitoring capability of cfDNA allows it to serve as a biomarker for efficacy evaluation. By tracking changes in cfDNA methylation profiles after intervention, clinicians can intuitively judge therapeutic response, enabling dynamic and precise management of CVD diagnosis and treatment. In addition, both RF and XGBoost models in this study identified age as the primary predictor of cardiac dysfunction risk, which is consistent with clinical experience. Aging itself is associated with epigenetic dysregulation, and superimposed degenerative changes in the cardiovascular system further aggravate epigenetic abnormalities in cardiomyocytes, ultimately increasing the risk of cardiac dysfunction. These findings suggest that in clinical application, cfDNA epigenetic results should be interpreted in combination with patient age.

Several limitations of this study should also be acknowledged. First, this is a single-center retrospective study with inherent selection bias. A future multicenter, prospective study with larger sample size is warranted. Second, Locus-specific methylation analysis and functional enrichment studies were not performed in this study. Third, Subtype-specific analysis was not feasible due to limited sample size; future studies should include larger cohorts of specific CVD subtypes. Fourth, the dynamic changes of cfDNA epigenetic characteristics and their associations with long-term adverse outcomes remain to be explored.

Conclusion

The epigenetic characteristics of cfDNA are closely associated with the severity of cardiac function in CVD patients and may serve as effective noninvasive molecular markers for cardiac function stratification. The prediction model integrating cfDNA epigenetic typing with traditional clinical indicators significantly improves the accuracy of risk prediction. Future studies should focus on locus-specific methylation, dynamic monitoring of cfDNA, and multicenter prospective validation to facilitate the translation of these findings into clinical practice.

Disclosure of conflict of interest

None.

Address correspondence to: Chunwen Jia, Department of Cardiovascular Medicine, Zhongshan

Hospital Affiliated to Xiamen University, No. 201-209, Hubin South Road, Siming District, Xiamen 361000, Fujian, China. E-mail: jiachunwen@163.com

References

[1] Armocida B, Formenti B, Silano M, Aszalos A, Bueno H, Elyamani M, Zabala AF, Leonardi M, Lindström J, Michalak K, Morillo B, Nowicki M, Oprešnik D, Palmieri L, Papa R, Peltonen M, Safadi H, Shchetynina H, Skogberg N, Tolonen H, Wouters E, Zaletel J and Onder G; JACARDI Consortium. Tackling the challenge of cardiovascular diseases and diabetes across Europe: a joint action by more than 300 public health professionals. Commentary. Ann Ist Super Sanita 2024; 60: 4-7.

[2] Lv Y, Zou Z, Chen GM, Jia HX, Zhong J and Fang WW. Amlodipine and angiotensin-converting enzyme inhibitor combination versus amlodipine monotherapy in hypertension: a meta-analysis of randomized controlled trials. Blood Press Monit 2010; 15: 195-204.

[3] Giannakopoulou SP, Antonopoulos A and Panagiotakos D. Serum inflammatory markers used in cardiovascular disease risk prediction models: a systematic review. Angiology 2026; 77: 137-144.

[4] Barraclough JY, Gandomkar Z, Fletcher RA, Barbieri S, Kuo NI, Rodgers A, Douglas K, Poppe KK, Woodward M, Luxan BG, Neal B, Jorm L, Brennan P and Arnott C. Predicting cardiovascular events from routine mammograms using machine learning. Heart 2026; 112: 261-269.

[5] Delpino FM, Costa ÂK, Farias SR, Chiavegatto Filho ADP, Arcêncio RA and Nunes BP. Machine learning for predicting chronic diseases: a systematic review. Public Health 2022; 205: 14-25.

[6] Yao H, Golbus JR, Gryak J, Pagani FD, Aaronson KD and Najarian K. Identifying potential candidates for advanced heart failure therapies using an interpretable machine learning algorithm. J Heart Lung Transplant 2022; 41: 1781-1789.

[7] Wang JL, Guo CY, Li HW, Zhao XQ and Zhao SM. Prognostic value of NT-proBNP in patients with successful PCI for ACS and normal left ventricular ejection fraction. Am J Med Sci 2022; 363: 333-341.

[8] Lo YMD, Han DSC, Jiang P and Chiu RWK. Epigenetics, fragmentomics, and topology of cell-free DNA in liquid biopsies. Science 2021; 372: eaaw3616.

[9] Zhang K, Fu R, Liu R and Su Z. Circulating cell-free DNA-based multi-cancer early detection. Trends Cancer 2024; 10: 161-174.

[10] Qi T, Pan M, Shi H, Wang L, Bai Y and Ge Q. Cell-free DNA fragmentomics: the novel promising biomarker. Int J Mol Sci 2023; 24: 1503.

[11] Loy C, Ahmann L, De Vlaminck I and Gu W. Liquid biopsy based on cell-free DNA and RNA. Annu Rev Biomed Eng 2024; 26: 169-195.

[12] Stawski R, Nowak D and Perdas E. Cell-free DNA: potential application in COVID-19 diagnostics and management. Viruses 2022; 14: 321.

[13] Schroers-Martin JG and Alizadeh AA. Cell-free DNA in hematologic malignancies. JCO Oncol Pract 2024; 20: 1491-1499.

[14] Edwards RL, Menteer J, Lestz RM anf Baxter-Lowe LA. Cell-free DNA as a solid-organ transplant biomarker: technologies and approaches. Biomark Med 2022; 16: 401-415.

[15] Thorsen SU, Moseholm KF and Clausen FB. Circulating cell-free DNA and its association with cardiovascular disease: what we know and future perspectives. Curr Opin Lipidol 2024; 35: 14-19.

[16] Artner T, Sharma S and Lang IM. Nucleic acid liquid biopsies in cardiovascular disease: cell-free DNA liquid biopsies in cardiovascular disease. Atherosclerosis 2024; 398: 118583.

[17] Liu Y. At the dawn: cell-free DNA fragmentomics and gene regulation. Br J Cancer 2022; 126: 379-390.

[18] Dworzynski K, Roberts E, Ludman A and Mant J; Guideline Development Group of the National Institute for Health and Care Excellence. Diagnosing and managing acute heart failure in adults: summary of NICE guidance. BMJ 2014; 349: g5695.

[19] Rosamond WD, Chang PP, Baggett C, Johnson A, Bertoni AG, Shahar E, Deswal A, Heiss G and Chambless LE. Classification of heart failure in the atherosclerosis risk in communities (ARIC) study: a comparison of diagnostic criteria. Circ Heart Fail 2012; 5: 152-9.

[20] Shi Y, Zhang H, Huang S, Yin L, Wang F, Luo P and Huang H. Epigenetic regulation in cardiovascular disease: mechanisms and advances in clinical trials. Signal Transduct Target Ther 2022; 7: 200.

[21] Tang Z, Yu Y, Ng K, Sow D, Hu J and Mei J. Disease network delineates the disease progression profile of cardiovascular diseases. J Biomed Inform 2021; 115: 103686.

[22] Glezeva N, Moran B, Collier P, Moravec CS, Phelan D, Donnellan E, Russell-Hallinan A, O'Connor DP, Gallagher WM, Gallagher J, McDonald K, Ledwidge M, Baugh J, Das S and Watson CJ. Targeted DNA methylation profiling of human cardiac tissue reveals novel epigenetic traits and gene deregulation across different heart failure patient subtypes. Circ Heart Fail 2019; 12: e005765.

[23] Raphael C, Briscoe C, Davies J, Ian Whinnett Z, Manisty C, Sutton R, Mayet J and Francis DP. Limitations of the New York Heart Association functional classification system and self-reported walking distances in chronic heart failure. Heart 2007; 93: 476-82.

[24] Vancheri F, Longo G and Henein MY. Left ventricular ejection fraction: clinical, pathophysiological, and technical limitations. Front Cardiovasc Med 2024; 11: 1340708.

Table S1. Distribution of CVD subtypes between mild and severe cardiac insufficiency groups

CVD subtype	Mild group (n=437)	Severe group (n=187)	*P* value*
Coronary heart disease [n (%)]	172 (39.4%)	92 (49.2%)	0.018
Atrial fibrillation [n (%)]	89 (20.4%)	63 (33.7%)	0.005

P* value calculated by chi-square test for comparison between mild and severe groups. Footnote: Data for coronary heart disease and atrial fibrillation were extracted from the baseline characteristics (Table 1**). Other CVD subtypes (e.g., cardiomyopathy, valvular heart disease, heart failure with preserved ejection fraction) were not separately reported due to limited case numbers and lack of detailed subtype classification in the original dataset. The distribution of these subtypes was balanced between the two groups (data not shown). The definitions of CVD subtypes were based on the ESC/ACC diagnostic criteria, as described in the Methods section.