International Journal of Clinical Medicine & Pharmacology

https://doi.org/10.70731/g7k4ek15

Associations Between Urinary and Blood Heavy Metal Exposure and Heart Failure Risk in Elderly Adults: Insights From an Interpretable Machine Learning Model Based on NHANES (2003-2020)

Yuting Yang [a,b], Shan Deng [a,b,c,d,*],

[a] *Department of Cardiology, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China.*
[b] *Clinic Center of Human Gene Research, Union Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, 430022, China.*
[c] *Hubei Key Laboratory of Metabolic Abnormalities and Vascular Aging, Huazhong University of Science and Technology, Wuhan, China.*
[d] *Hubei clinical research center for metabolic and cardiovascular disease, Huazhong University of Science and Technology, Wuhan, China.*

KEYWORDS

Heavy Metals,
Heart Failure,
Machine Learning,
NHANES

ABSTRACT

This study examines the link between heavy metal exposure and heart failure risk in individuals aged 50 and older, utilizing machine learning models and data from the NHANES dataset. Five models were evaluated, with Gradient Boosting Decision Trees (GBDT) selected for its accuracy, interpretability, and ability to capture nonlinear relationships. The GBDT model achieved an accuracy of 0.78, sensitivity of 0.93, and an AUC of 0.92. Analysis revealed that higher levels of urinary iodine, blood cadmium, urinary cobalt, urinary tungsten, and urinary arsenic acid were significantly associated with increased heart failure risk. Synergistic effects with age and BMI further amplified these risks. Interpretability tools such as SHAP, permuted Feature Importance, ICE, and PDP were used to enhance model transparency and understanding. These findings underscore the importance of ongoing research into the mechanisms linking heavy metals and heart failure and the need for monitoring and regulatory measures to protect vulnerable populations.

1. Introduction

Individuals are unavoidably exposed to environments containing heavy metals, and these elements can infiltrate the human body via ingestion and inhalation[1]. The threat of heavy metals to human health continues to exist and has now become a global public health issue. Heavy metals typically function by displacing essential metal ions in the human body[2], leading to disruptions in critical biological pathways. Such disruptions can result in endothelial cell damage[3-5], disturbances in lipid metabolism[6], increased oxidative stress[7, 8], immune

* Corresponding author. E-mail address: dengshan1020@qq.com

Received 18 February 2025; Received in revised form 23 February 2025; Accepted 26 February 2025; Available online 28 February 2025.

homeostasis imbalance[9, 10], and epigenetic modifications[11], all of which may contribute to the development of various cardiovascular diseases. Exposure to heavy metals during pregnancy has been linked to an increased risk of congenital heart defects in offspring [12, 13]. Additionally, exposure in preschool children may lead to arrhythmias[14], while heavy metal exposure has been associated with hypertension[15]. The impact of exposure to some heavy metals on the incidence of heart failure has also been studied[16, 17].

Heart failure isn't a standalone illness but rather the end stage of various heart diseases, which continues to be a significant contributor to mortality, morbidity, and diminished quality of life globally[18]. Heart failure impacts the health of more than 60 million people worldwide[19], and the lifetime risk of heart failure has risen to 24%, roughly one in four people will develop heart failure in their lifetime[20]. Research indicates a significant correlation between cadmium exposure and the incidence of heart failure[21-24]. And the level of urinary antimony appears to be directly proportional to the risk of heart failure[25]. Besides, epidemiological research has indicated that exposure to cobalt and lead elevates the risk of heart failure[16, 26]. However, previous research has predominantly relied on conventional statistical methods, which exhibit limited flexibility in modeling complex relationships and often operate under restrictive assumptions regarding underlying data distributions[27]. Furthermore, much of the existing literature has focused on examining the effects of a single type of heavy metal in isolation. This approach may overlook potential interactions and cumulative effects of multiple heavy metals, thereby limiting the comprehensiveness of the findings. Thus, a more integrated and flexible modeling framework, machine learning, may be necessary to capture the intricacies of heavy metal exposure and its multifaceted impact on heart failure risk.

Our study conducted a comprehensive analysis of laboratory indicators for various heavy metal elements in both blood and urine, utilizing data from the National Health and Nutrition Examination Survey (NHANES) database spanning 2003 to 2020. This robust database provides high-quality, nationally representative data essential for public health research, integrating interviews, physical examinations, and laboratory assessments. Furthermore, we employed advanced machine learning algorithms for the analysis, enhancing our ability to uncover meaningful relationships within the data. Machine learning tends to prioritize predictive performance and generalization over interpretability, with a mathematical emphasis on cross-validation and iterative enhancement of algorithms[28]. This implies that machine learning algorithms frequently outshine traditional statistical prediction methods, however, less transparent compared to conventional statistical approaches. Consequently, there is a need for explanatory models in machine learning. SHAP (SHapley Additive exPlanations) is a technique for interpreting machine learning models, employing the computation of Shapley values to ascertain the contribution of each feature to a given prediction, and thus facilitates the generation of transparent explanations by leveraging these calculated Shapley values[29, 30]. In this study, we chose five machine learning algorithms for our analysis, identified the best-performing model, and utilized the SHAP model for interpreting.

2. Method

2.1. Data Source

All variable information was sourced from the NHANES (the National Health and Nutrition Examination Survey) database (https://wwwn.cdc.gov/Nchs/Nhanes). The NHANES database, carried out by the Centers for Disease Control and Prevention (CDC) in the United States, combines interviews, laboratory tests and physical examinations on a nationally representative sample of American residents every two years. We downloaded the variables data from the annual datasets of NHANES 2003-2004, NHANES 2005-2006, NHANES 2007-2008, NHANES 2009-2010, NHANES 2011-2012, NHANES 2013-2014, NHANES 2015-2016, NHANES 2017-2018, and NHANES 2019-2020, match all variables for each year by the unique sequence numbers, and ultimately merged the datasets for all the years. The features and labels encompassed in the study include gender, age (year), race, education level, poverty-to-income ratio (PIR), height, weight, body mass index (BMI, kg/m^2), urinary heavy metals (including urine mercury, urine barium, urine beryllium, urine cadmium, urine cobalt, urine cesium, urine molybdenum, urine lead, urine platinum, urine antimony, urine thallium, urine tungsten, urine uranium, urine arsenous acid, urine arsenobetaine, urine arsenocholine, urine dimethylarsonic acid, urine monomethylacrsonic acid, urine total arsenic, and

urine iodine), blood heavy metals (including blood cadmium, blood lead, blood mercury), and heart failure. The diagnosis of heart failure is ascertained through self-reported physician diagnoses acquired via standardized medical questionnaires in individual interviews. Participants were queried: "Has a doctor or other health professional ever told that you had congestive heart failure?" If a person responds 'yes', they will be regarded as having heart failure. According to the 10th revised edition of the International Classification of Diseases and Related Health Problems (ICD-10), the code for congestive heart failure is I50.0, I50.1, I50.9[31]. Heavy metal samples in blood and urine were processed, stored, and transported to the Centers for Disease Control and Prevention and the Science Department of the Environmental Health Laboratory of the National Environmental Health Center for analysis.

2.2. Study Population

The inclusion criteria were as follows: (1) participants were aged 50 or above; (2) participants underwent heavy metal laboratory testing in blood and urine samples; (3) heart failure status was 1 (ever had a heart failure) or 2 (didn't have a heart failure) based on the NHANES questionnaires. Exclusion criteria were as follows: (1) heart failure status was 7 (refused to answer this question) or 9 (means uncertain heart failure status) according to NHANES questionnaires. Finally, a total of 6879 participants were included in the present analysis.

2.3. Statistical Analysis

2.3.1. Statistical Description of Baseline Characteristics Between Heart Failure Group and no Heart Failure Group

The demographic description section was analyzed using R 4.2.3. In the baseline table, we divided the participants into the heart failure group and the no heart failure group for comparison. The baseline covariates were presented as mean (standard deviation) for continuous variables and count (percentage) for categorical variables, respectively. The t tests and $\chi 2$ tests were employed to compare the differences between the groups for continuous and categorical variables, separately. variables associated with heavy metals concentration were depicted using geometric mean and geometric standard deviation.

2.3.2. Machine Learning Processing

2.3.2.1. Data Preprocessing and Splitting

The analysis of the machine learning section is accomplished using Python 3.10.8 and Python libraries, including "sklearn", "imblearn", "numpy", and "pandas". An overview of our methodology is presented in (Figure 1). BMI, age, gender, blood heavy metals, and urinary heavy metals were selected as feature variables, and heart failure as the target variable for the machine learning model. "missForest" was used for missing value imputation, which is an iterative imputation approach based on random forest method and can concurrently interpolate different types of data[32]. The raw data was standardized using the "StandardScaler" function. The dataset was randomly split into a training set (N=5159) and a testing set (N=1720) in accordance with a specific ratio, with the training set constituting 75% of all the dataset. We calculated the Pearson correlation coefficients of all predictor variable in the training set and plotted heatmaps respectively before and after deleting highly correlated variables (Supplementary figure 1). To reduce the influence of multicollinearity on the model, features with Pearson coefficients exceeding 0.8 were removed[33]. The low variance features have been removed, which possess limited explanatory information and make a minor contribution to the model's predictive capacity.

2.3.2.2. Feature Selection and Model Training

After the initial feature filtering via the above-mentioned steps, we incorporated a pipeline consisting of "RandomUnderSample", "SelectKBest", and the model classifier for the subsequent analysis. Among the pipeline process, RUS (RandomUnderSample) is a sampling method that can create balanced classes from imbalanced ones[34, 35], and "SelectKBest" is a feature selection method that functions by scoring all features and selecting the top "k" input variables with the highest scores[36] (mutual information was selected as the scoring metric, with higher scores indicating a stronger correlation between the variables[37]). Grid search was used for hyperparameter tuning, thereby selecting the optimal hyperparameter values for each machine learning model[38] (The detailed parameter information for the model is provided in the supplementary materials). We chose five machine learning models: SVM (Support Vector Machine), RF (Random Forest), XGBoost (eXtreme Gradient Boosting), GDBT (Gradient Boosting Deci-

sion Tree), and KNN (k-nearest neighbor). Among them, the GDBT was found to be the best one.

2.3.2.3.Model Evaluation and Comparison

To conduct a thorough and systematic comparison of the machine learning models, we employed a multifaceted approach that included generating various evaluation plots alongside calculating a comprehensive set of performance metrics both in the training and testing sets. Specifically, we constructed Receiver Operating Characteristic (ROC) curves, precision-recall (PR) curves, and confusion matrices. The ROC curves provide insight into the model's diagnostic capabilities and the PR curves offer a focused analysis of the model's performance in classifying positive cases, particularly in imbalanced datasets. Confusion matrices were used to provide a detailed breakdown of the models' classification results, highlighting the number of true positives, true negatives, false positives, and false negatives. We further calculated a comprehensive suite of quantitative metrics, including accuracy, specificity, precision, sensitivity (recall), negative predictive value (NPV), positive predictive value (PPV), false positive rate (FPR), false negative rate (FNR), false discovery rate (FDR), the F1 score, and the Brier score. These metrics collectively elucidated the models' classification efficacy, with the F1 score serving to balance precision and recall and the Brier score reflecting the probabilistic accuracy of predictions. Moreover, we evaluated the area under the ROC curve (AUC) and its 95% confidence intervals as a measure of the models' overall ability to distinguish between positive and negative classes, complemented by the average precision score (AP) to quantify performance in scenarios characterized by class imbalance.

2.3.2.4.Model Explanation

In the explanation section of the optimal model, we calculated the PI (permutation importance) values of each feature in the model, constructed the SHAP (SHapley Additive exPlanations) explanation model using the "shap" package[39], and plotted ICE (Individual Conditional Expectation) plot and PDP (Partial Dependence Plot) to respectively demonstrate the impact of a single feature and two features on the target.

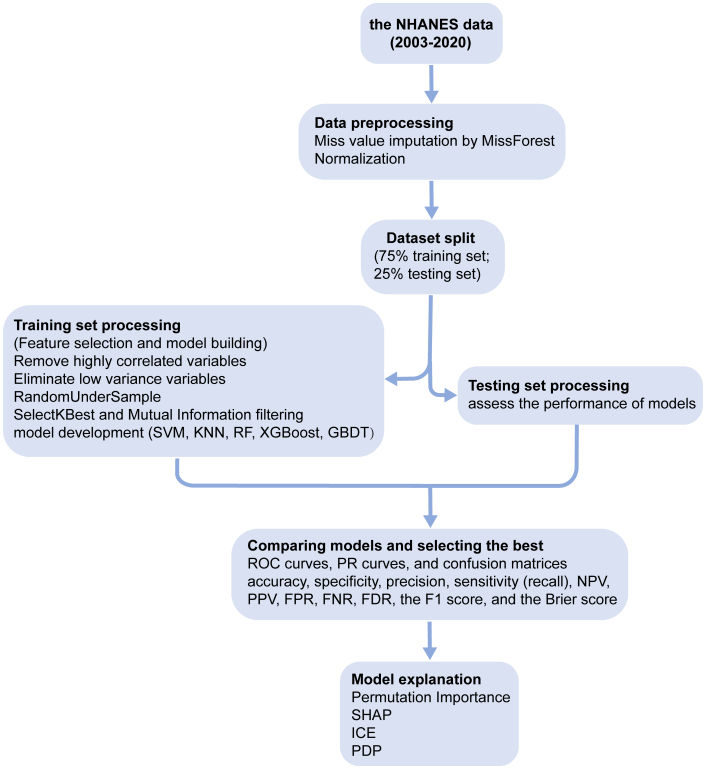


Figure 1 | Machine learning model development flowchart.

This flowchart outlines the key steps in developing a machine learning model from inception to interpretation.

3. Results

3.1. Baseline Characteristics

Among 6,879 participants, 388 were diagnosed with heart failure, while 6,491 were classified as having no heart failure. As illustrated in Table 1, a substantial proportion of the baseline characteristics exhibited significant statistical differences between the heart failure group and the group without heart failure. Compared to participants without heart failure, those in the heart failure group exhibited a higher age (69.74 $\pm$ 8.39 versus 65.33 $\pm$ 9.04; P<0.01), a greater proportion of male participants (60.1% versus 49.6%; P<0.01), and an elevated BMI (32.92 $\pm$ 16.51 versus 29.53 $\pm$ 7.22; P<0.01). Besides, significant statistical differences (P Value < 0.05) were observed in the laboratory indicators of heavy metals, including blood cadmium, blood lead, blood mercury, urinary beryllium, urinary cadmium, urinary cesium, urinary antimony, urinary thallium, urinary tungsten, urinary uranium, urinary dimethylarsonic acid, urinary total arsenic, and urinary iodine, between the two participant groups.

Table 1 | Baseline characteristics of study participants

Characteristics	Total adults (N= 6879)	Heart failure(N= 388)	No heart failure(N= 6491)	P value
Age, years, mean (SD)	65.58(9.06)	69.74(8.39)	65.33(9.04)	<0.01
Gender, n (%)				<0.01
Male	3454(50.2)	233(60.1)	3221(49.6)	
Female	3425(49.8)	155(39.9)	3270(50.4)	
BMI, kg/m^2, mean (SD)	29.72(8.07)	32.92(16.51)	29.53(7.22)	<0.01
Weight status, n (%)				
BMI <25	1727(25.1)	79(20.4)	1648(25.4)	<0.01
BMI ⩾25 & <30	2391(34.8)	108(27.8)	2283(35.2)	
BMI ⩾30	2761(40.1)	201(51.8)	2560(39.4)	
Race, n (%)				<0.01
Mexican American	812(11.8)	27(7)	785(12.1)	
Hispanic	714(10.4)	31(8)	683(10.5)	
Non-Hispanic White	2954(42.9)	196(50.5)	2758(42.5)	
Non-Hispanic Black	1613(23.4)	109(28.1)	1504(23.2)	
Other Race (Including Multi-Racial)	786(11.4)	25(6.4)	761(11.7)	
Education, n (%)				<0.01
Grades 0–12	1900(27.6)	118(30.4)	1782(27.5)	
High school graduate/GED or equivalent	3485(50.7)	223(57.5)	3262(50.3)	
College graduate or above	1494(21.7)	47(12.1)	1447(22.3)	
PIR, mean (SE)	2.61(1.56)	2.02(1.31)	2.65(1.56)	<0.01
PIR level, n (%)				<0.01
Low (<1)	1112(16.2)	83(21.4)	1029(15.9)	
Medium (⩾1 & <4)	4061(59)	257(66.2)	3804(58.6)	
High (⩾4)	1706(24.8)	48(12.4)	1658(25.5)	
Blood heavy metals, geometric mean (GSD)				
Cadmium (ug/L)	0.40(0.76)	0.48(0.79)	0.40(0.76)	<0.01
Lead (ug/dL)	1.47(0.62)	1.63(0.61)	1.46(0.62)	0.02
Mercury, total (ug/L)	0.95(1.01)	0.72(0.89)	0.96(1.02)	<0.01
Urinary heavy metals, geometric mean (GSD)				
Mercury (ug/L)	0.30(1.08)	0.25(1.05)	0.30(1.08)	0.05
Barium (ug/L)	0.96(1.03)	0.83(1.09)	0.97(1.02)	0.55
Beryllium (ug/L)	0.60(1.50)	0.55(1.41)	0.60(1.50)	<0.01
Cadmium (ug/L)	0.32(0.98)	0.37(0.98)	0.31(0.98)	<0.01
Cobalt (ug/L)	0.34(0.88)	0.43(0.91)	0.34(0.87)	0.15
Cesium (ug/L)	4.09(0.65)	3.63(0.60)	4.12(0.65)	<0.01
Molybdenum (ug/L)	34.86(0.87)	36.14(0.79)	34.79(0.70)	0.83
Lead (ug/L)	0.48(0.89)	0.49(0.94)	0.48(0.89)	0.49
Platinum (ug/L)	0.40(2.47)	0.38(2.38)	0.40(2.48)	0.08
Antimony (ug/L)	0.05(0.98)	0.07(1.22)	0.05(0.97)	<0.01
Thallium (ug/L)	0.15(0.81)	0.15(0.10)	0.15(0.80)	<0.01
Tungsten (ug/L)	0.06(1.14)	0.09(1.26)	0.06(1.13)	<0.01
Uranium (ug/L)	0.03(2.11)	0.04(2.31)	0.03(2.10)	<0.01
Arsenous acid (ug/L)	0.30(1.12)	0.30(1.16)	0.30(1.12)	0.1
Arsenic acid (ug/L)	0.64(0.24)	0.66(0.31)	0.63(0.24)	0.23
Arsenobetaine (ug/L)	2.51(1.54)	2.50(1.47)	2.51(1.54)	0.14
Arsenocholine (ug/L)	0.18(0.91)	0.20(0.99)	0.18(0.90)	0.21
Dimethylarsonic acid (ug/L)	3.63(0.81)	3.23(0.71)	3.66(0.82)	<0.01
Monomethylacrsonic acid (ug/L)	0.48(0.91)	0.50(0.89)	0.47(0.91)	0.86
Total arsenic (ug/L)	8.55(1.16)	7.80(1.06)	8.60(1.17)	0.03
Iodine (ng/mL)	146.76(0.95)	183.23(1.28)	144.83(0.92)	<0.01

All continuous demographic variables are presented as mean (standard deviation), while the categorical variables are presented as frequency (percentage). Variables related to heavy metals are represented using geometric mean (geometric standard deviation). Statistical comparisons between the heart failure and no heart failure groups were conducted using Student's t-test for continuous variables and chi-square tests for categorical variables, with a significance level set at $p < 0.05$. BMI: Body Mass Index is calculated as weight in kilograms divided by the square of height in meters; PIR: Poverty to Income Ratio is calculated based on household income relative to the federal poverty line.

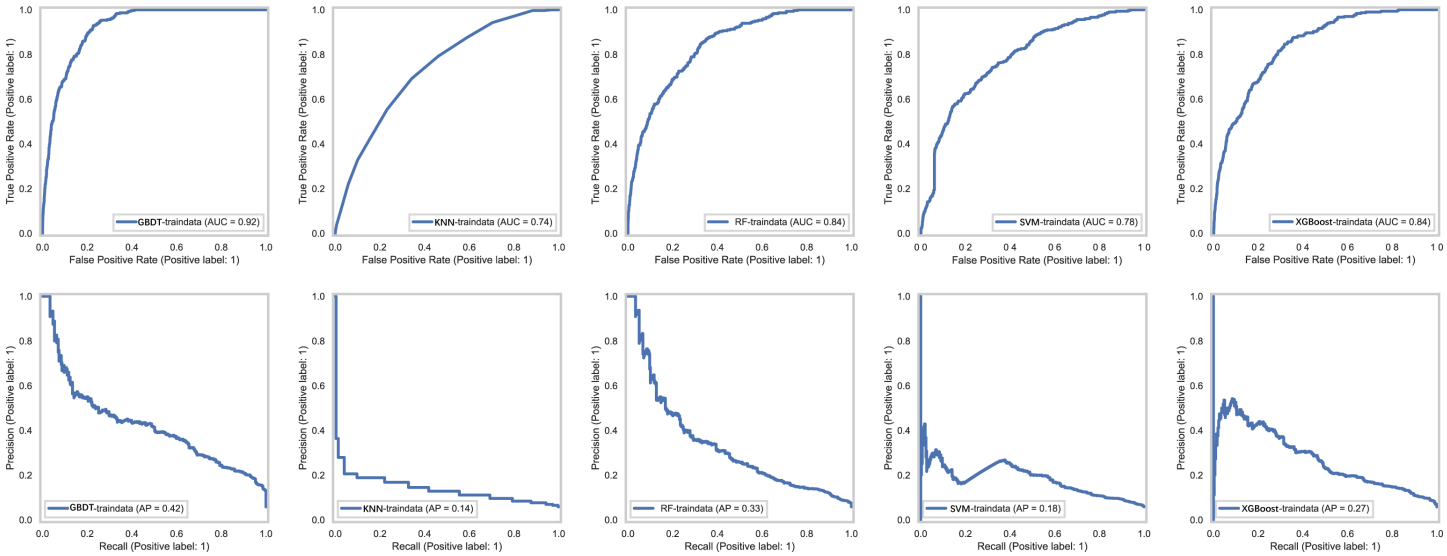


Figure 2 | ROC and PR curves of the five models in the training set

The figure depicts the ROC curves and PR curves for 5 predictive models: GBDT, KNN, RF, SVM, and XGBoost. Each model's performance is shown, demonstrating its ability to distinguish between the positive and negative classes (ROC) and the balance between precision and recall (PR).

Table 2 | Evaluation indicators for predictive models.

Model	Dataset	Accuracy	Sensitivity	Specificity	Precision	Recall	NPV	PPV	FPR	FNR	FDR	F1 score	Brier score	AUC	AUC_low	AUC_up	AP score
GBDT	Training	0.78	0.93	0.77	0.20	0.93	0.99	0.20	0.23	0.07	0.80	0.33	0.18	0.92	0.91	0.93	0.42
GBDT	Testing	0.75	0.71	0.75	0.14	0.71	0.98	0.14	0.25	0.29	0.86	0.23	0.19	0.81	0.77	0.85	0.17
KNN	Training	0.76	0.56	0.77	0.13	0.56	0.97	0.13	0.23	0.44	0.87	0.21	0.21	0.74	0.71	0.76	0.14
KNN	Testing	0.74	0.57	0.75	0.11	0.57	0.97	0.11	0.25	0.43	0.89	0.19	0.22	0.73	0.68	0.78	0.11
RF	Training	0.68	0.85	0.67	0.14	0.85	0.99	0.14	0.33	0.15	0.86	0.24	0.20	0.84	0.82	0.86	0.33
RF	Testing	0.67	0.70	0.67	0.10	0.70	0.98	0.10	0.33	0.30	0.90	0.18	0.21	0.77	0.73	0.82	0.14
SVM	Training	0.80	0.62	0.81	0.16	0.62	0.97	0.16	0.19	0.38	0.84	0.26	0.21	0.78	0.76	0.81	0.18
SVM	Testing	0.77	0.58	0.78	0.13	0.58	0.97	0.13	0.22	0.42	0.87	0.21	0.22	0.75	0.69	0.80	0.12
XGBoost	Training	0.69	0.84	0.68	0.14	0.84	0.99	0.14	0.32	0.16	0.86	0.24	0.21	0.84	0.82	0.86	0.27
XGBoost	Testing	0.68	0.65	0.68	0.10	0.65	0.97	0.10	0.32	0.35	0.90	0.17	0.21	0.75	0.70	0.80	0.15

NPV: Negative Predictive Value. FPR: False Positive Rate. FNR: False Negative Rate. FDR: False Discovery Rate. F1 score: The harmonic mean of precision and recall, calculated as 2 * (Precision * Recall) / (Precision + Recall). AUC: Area Under the Curve. AP: Average Precision.

3.2. Comparison of Machine Learning Models

Considering our dataset is imbalanced, performance metrics that account for class distribution, such as sensitivity (recall), precision, F1 score, and AUC, become more crucial in determining which model is the best, compared with other evaluating indicators. Figure 2 illustrates the ROC curves and PR curves for the five machine learning models for the training set, presented separately. The AUC for GBDT was 0.92 (95% CI: 0.91-0.93), indicating superior predictive performance in distinguishing heart failure compared to the four other models. The RF yielded an AUC of 0.84 (95% CI: 0.82-0.86), which was similar to that of XGBoost, at 0.84 (95% CI: 0.82-0.86), suggesting robust performance. In contrast, the SVM achieved an AUC of 0.78 (95% CI: 0.76-0.81), while the K-Nearest Neighbors (KNN) model had an AUC of 0.74 (95% CI: 0.71-0.76), indicating a moderate discrimination ability. Besides, GBDT stand out with the highest average precision score, demonstrating superior capability in maintaining precision and better diagnostic accuracy in heart failure compared to other models. Table 2 presents specific evaluation metrics for five machine learning models. GBDT achieves a sensitivity of 0.93, indicating excellent true positive identification. In contrast, KNN and SVM exhibit considerably lower sensitivities of 0.56 and 0.61, respectively, resulting in a greater likelihood of missing actual heart failure cases. Overall, GBDT demonstrates distinct advantages over KNN, RF, SVM, and XGBoost based on these performance metrics. Its superior sensitivity (0.93), Average Precision Score (APS) of 0.42, and accuracy of 0.78 position GBDT as an exceptional choice for predictive modeling in scenarios where both precision and recall are essential. The ROC curves and PR curves for the testing

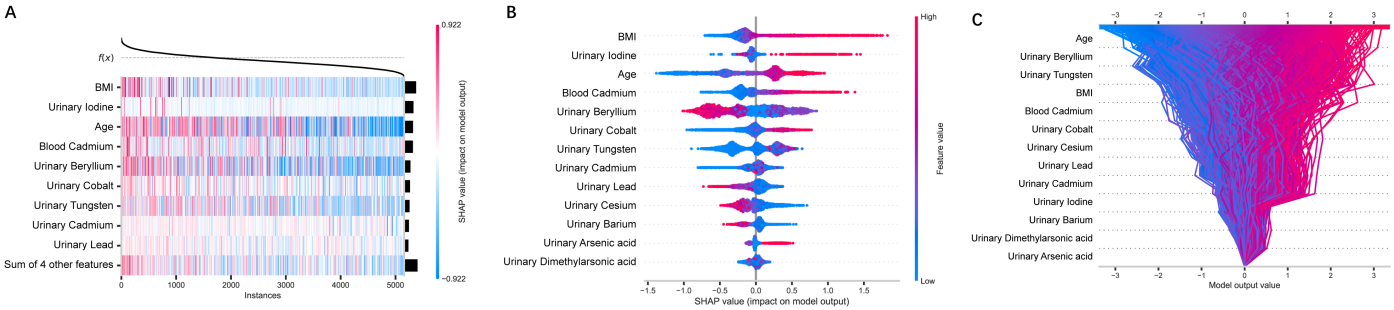


Figure 3 | SHAP visualizations for model interpretation

Figure 3A: SHAP heatmap. The y-axis displays different features arranged in descending order based on their maximum absolute SHAP values, indicating their relative contributions to the prediction of heart failure outcomes. The x-axis represents individual observations from the dataset. Each cell in the heatmap is color-coded to reflect the SHAP values, with red indicating positive contributions to heart failure risk and blue representing negative contributions. Figure 3B: SHAP beeswarm map. The features on the y-axis are organized in the same order as those in the heatmap. The x-axis quantifies the SHAP values, reflecting the impact of each feature on the model's predictions. Each point on the plot corresponds to a specific observation, with colors representing feature value levels; red indicates higher levels while blue denotes lower levels. Figure 3C: SHAP decision map. Illustrates individual predictions in relation to the cumulative effects of features, showing how each feature contributes to the final decision. Blue indicates negative predictions, while red indicates positive predictions.

set are presented in supplementary figure 2, and supplementary figure 3 displays the corresponding confusion matrices.

3.3. The Explainability of the Best Machine Learning Model

3.3.1. SHAP Visualization Plot

Figure 3A presents a SHAP heatmap that offers a comprehensive visualization of the contributions of various features to the predictions made by GBDT regarding heart failure risk. The model's output is shown above the heatmap, centered around the expected SHAP values for all observations. Additionally, the global importance of each input feature is depicted as a bar plot on the right side of the heatmap. Notably, BMI, urinary iodine, age, and blood cadmium emerge significant contributors to the model's predictions. Figure 3B presents the SHAP beeswarm plot for GBDT, illustrating the relationship between feature values and the risk of heart failure. Features such as urinary iodine, blood cadmium, urinary cobalt, body mass index (BMI), and age occupy the top positions and exhibit significant positive SHAP values, suggesting a strong association with an increased risk of heart failure. Urinary tungsten, urinary cadmium, and urinary arsenic acid are positioned further down the y-axis with lower SHAP value, indicating a minimal positive contribution to the risk of heart failure. Interestingly, lower levels of urinary beryllium, urinary lead, urinary cesium, and urinary barium are associated with higher SHAP values, suggesting a potential protective effect against heart failure. The SHAP decision plot for GBDT is illustrated in Figure 3C. Each line represents an individual participant, with the trajectory of each line converging from the bottom and extending toward the top indicating how the features influence the model's output for specific instances. Figure 4 depicts the SHAP interaction summary plot, where the horizontal axis represents interaction values that quantify the attribution of paired interaction effects between two features. Most variables demonstrate varying degrees of interaction effects. Features such as age, urinary beryllium, urinary tungsten, and BMI exhibit significant interaction effects with other variables, indicating the complex relationships that influence heart failure risk.

3.3.2. Permutation Importance Analysis

The permutation importance analysis was conducted to estimate the contribution of features to the prediction of heart failure risk within GBDT by measuring the decrease in the model's performance when the values of a feature were randomly shuffled (Figure 5). The results revealed that urinary beryllium and blood cadmium exhibited the highest importance scores of 0.147 and 0.084 among the heavy metal variables, indicating their significant influence on predictive accuracy. Age and BMI were identified as important determinants, with scores of 0.123 and 0.108, respectively.

3.3.3. ICE Plot and PDP Plot

The ICE plots (Figure 6), generated through partial conditional expectation, illustrated that heart failure risk increases with rising levels of age, BMI, and

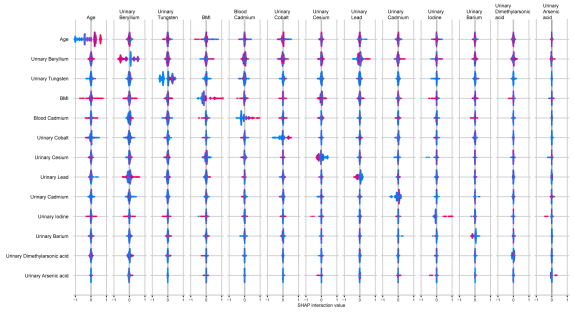
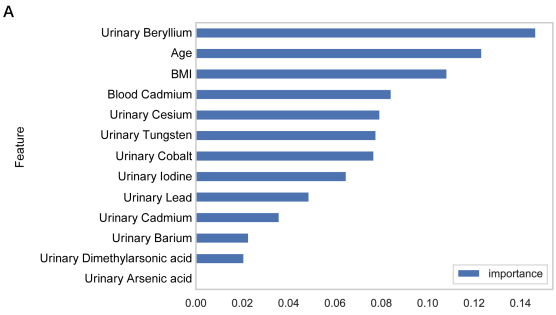


Figure 4 l SHAP interaction plot

Illustrate the combined effects of two features on model predictions.



B

Feature	Importance
Urinary Beryllium	0.147
Age	0.123
BMI	0.108
Blood Cadmium	0.084
Urinary Cesium	0.079
Urinary Tungsten	0.078
Urinary Cobalt	0.077
Urinary Iodine	0.065
Urinary Lead	0.049
Urinary Cadmium	0.036
Urinary Barium	0.023
Urinary Dimethylarsonic acid	0.021
Urinary Arsenic acid	0.000

Figure 5 l Premutation importance analysis

The figure evaluates the contribution of individual features to the model's predictive performance through permutation importance analysis. Figure 5A: Bar plot of feature importance scores, ranking the features according to their significance. Figure 5B: Table displaying the feature importance scores.

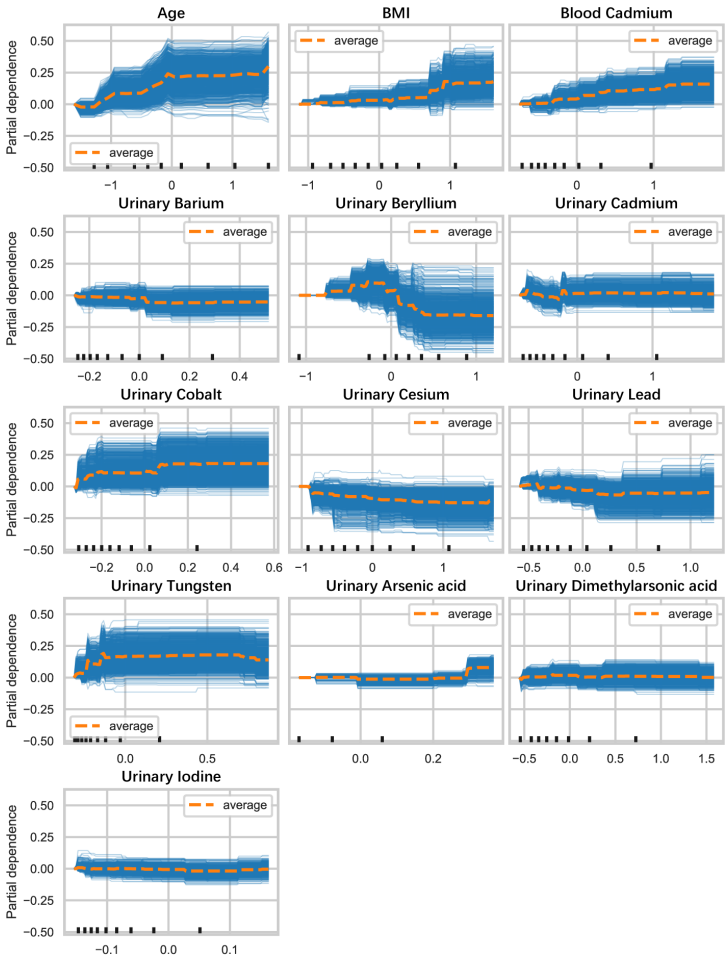


Figure 6 l Individual conditional Expectation (ICE) plots

This figure presents the ICE diagrams for each feature, enabling the examination of the effect of a single variable on predictions while eliminating the noise introduced by interactions with other features. Each thin blue line in the plot represents an individual observation, showcasing how the predicted response varies as the feature value changes, and each orange dashed line represents the average level.

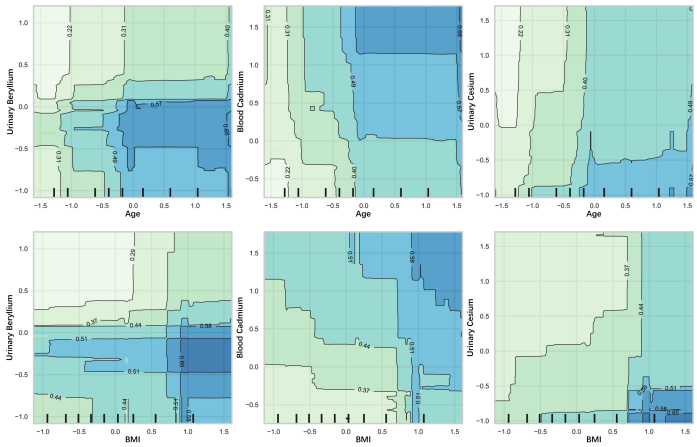


Figure 7 l Two-Dimensional Partial Dependence Plot (PDP)

This plot offers a comprehensive visualization of the interactive influence of age and BMI on the predicted outcomes associated with the top three metals ranked by importance. Lighter shades represent lower predicted outcomes, while darker shades indicate higher predictions. The equipotential lines illustrate levels of constant predicted outcomes across the combinations of the two features, signifying that any point along a given line produces the same average predicted response from the model.

blood cadmium, highlighting their potential as critical risk factors in assessments of heart failure. The relationship between urinary beryllium and heart failure risk was characterized by an inverted U-shaped curve, suggesting that risk is heightened within a specific range of exposure. Other assessed features displayed moderate effects, indicating their contribution to prediction but with less pronounced impacts. Additionally, the two-dimensional PDP plot clarified the joint predictive effects of the top three heavy metal variables in the permutation importance analysis, alongside age and gender, on heart failure risk

(Figure 7). The color gradient ranges from deep blue, indicating higher risk, to light green, reflecting lower risk. Notably, blood cadmium interacted synergistically with age and gender, suggesting potential combined effects on heart failure risk assessment.

4. Discussion

This study employed five machine learning models to explore the complex relationship between heavy metal exposure in blood and urine and the risk of heart failure. Following a thorough evaluation of various model performance metrics, the GBDT model was chosen as the optimal predictive model due to its superior accuracy and interpretability. To further explain the findings derived from GBDT, several machine learning interpretability techniques were applied, including SHAP (SHapley Additive exPlanations), permuted Feature Importance, individual conditional expectation (ICE), and partial dependence plots (PDP). These methods enabled a comprehensive assessment of the individual contributions of each feature to heart failure risk, as well as interactions between pairs of features, offering valuable insights into the prediction on heart failure. Notably, the interplay between different heavy metals and other demographic factors, such as age and BMI, suggests that risk assessments for heart failure must consider these complex interactions. Understanding these relationships not only furthers our knowledge of environmental determinants of cardiovascular disease but also emphasizes the need for continuous monitoring and potential regulatory measures regarding heavy metal exposure.

GBDT, the best prediction machine learning model in our study, is an ensemble learning technique that sequentially builds multiple decision trees, with each tree trained to correct the errors of its predecessor. The iterative approach allows GBDT to progressively minimize the residual errors through a method known as gradient descent[40, 41]. GBDT exhibits robustness against overfitting and has superior predictive accuracy compared to traditional models[41]. The capacity of GBDT to capture complex nonlinear relationships within the data is particularly pertinent in the context of researches where the relationships between features and targets are intricate, such as our study about the complex relationship between heavy metals and the risk of heart failure. In this study, we specifically focused on participants aged 50 years and older within the NHANES database. This demographic selection is particularly important given that older adults not only experience higher morbidity rates[42] but also significantly elevated mortality risks associated with heart failure[43]. Our analysis predominantly centered on heavy metals detected in blood and urine, underscoring our intent to investigate the impact of these environmental exposures on heart failure, intentionally setting aside other prevalent cardiovascular risk factors. And considering that our primary goal was to develop a robust predictive model capable of identifying at-risk individuals based on their exposure to heavy metals, we prioritized predictive accuracy over population-level inferences and chose not to apply sample weights in our analysis. Moreover, the NHANES database provides a large and diverse sample size conducive to model construction, alleviating concerns regarding the need for weighting and allowing us to derive meaningful insights from the unweighted data while maintaining a clear focus on predictive performance.

Our findings complement existing studies and, in certain instances, corroborate prior researches. Specifically, we found that blood cadmium is a positive predictor of heart failure, supporting previous findings. Major sources of cadmium exposure in the population include smoking and passive smoking[44, 45]. A case-cohort study reported that higher urinary cadmium levels are associated with an increased overall incidence of heart failure, with a hazard ratio (HR) of 1.1 per interquartile range difference (95% CI: 1.0-1.2)[17]. Furthermore, blood cadmium has been identified as an independent risk factor for heart failure, demonstrating an OR of 1.345 ($p < 0.001$)[46]. Similarly, an association study revealed a quantitative correlation between cadmium levels and heart failure risk, showing that a 50% increase in blood cadmium corresponds to a 48% increase in heart failure risk (OR: 1.48; 95% CI: 1.17-1.87), while a similar increase in urinary cadmium is associated with a 12% increase in risk (OR: 1.12; 95% CI: 1.03-1.20)[21]. The relationship between beryllium and lung cancer has been extensively researched[47], but its effects on cardiovascular health are less understood. Emerging studies suggest that beryllium exposure may increase the risk of cardiovascular disease[48], with significant associations observed in individuals with close contact to the metal, who demonstrate a heightened risk of ischemic heart disease mortality[49]. Cobalt, often found in human joint replacement implants, has also garnered attention due to its health implications[50]. Our research highlights

that cobalt exposure is linked to an increased risk of heart failure. A controlled trial indicated that patients undergoing multiple hip replacement surgeries exhibited significantly elevated myocardial cobalt levels, with autopsy reports revealing a prevalence of dilated heart chambers and decreased ejection fraction among these individuals[51]. In addition, we found that tungsten and arsenic also elevate the risk of heart failure. While both heavy metals are prevalent in the environment, research on their cardiovascular effects remains limited. A cross-sectional study indicated a positive correlation between urinary tungsten and cardiovascular disease risk[52, 53], and arsenic acid poisoning has been associated with significant alterations in heart rate, QRS duration, and symptoms of prolonged QT and QTc intervals due to arsenic acid poisoning[54]. In our study, we explored the combined effects of heavy metals, body mass index (BMI), and age on heart failure risk. Previous investigations have shown that heavy metals can significantly increase abdominal circumference and obesity rates, which may explain the observed synergistic effects of BMI and metals such as cadmium[55-57]. Additionally, there is a documented correlation between age and the duration of heavy metal exposure [58], and excessive accumulation of heavy metals in the human body may increase the risk of PhenoAge acceleration[59].

The value of our research lies in its multifaced approach, leveraging advanced computational techniques to enhance the understanding of environmental risk factors associated with heart failure. By integrating machine learning, we move beyond traditional statistical methods, enabling the identification of complex, nonlinear relationships. Furthermore, the significance of our findings contributes to the growing evidence linking heavy metals to adverse cardiovascular outcomes, highlighting potentially modifiable risk factors for clinical practice and public health strategies.

5. Limitation

While our study offers valuable insights into the relationship between heavy metal exposure and heart failure, several limitations warrant consideration. First, despite the GBDT model achieving an accuracy of 0.78, a high sensitivity of 0.93, and a strong AUC of 0.92, the model's relatively low precision and FDR indicate a need for refinement in its predictive capabilities. This suggests that the model may generate a considerable number of false positives, which can impact clinical decision-making. Second, while the use of machine learning algorithms has improved predictive accuracy compared to traditional methods, their inherent complexity may limit interpretability, which could impede acceptance in practical applications. Third, the cross-sectional design of the study restricts our ability to infer causality. While the observed associations are significant, they do not establish a direct cause-and-effect relationship between heavy metal exposure and heart failure outcomes.

6. Conclusion

In this study, GBDT demonstrated superior performance in modeling the relationship between heavy metal exposure and heart failure. Our analysis revealed that elevated concentrations of urinary iodine, blood cadmium, urinary cobalt, urinary tungsten, and urinary arsenic acid are significantly correlated with an increased risk of heart failure, while urinary beryllium appears to have a potentially detrimental effect. Additionally, we uncovered a synergistic relationship involving both age and BMI that amplifies the adverse effects of heavy metal exposure on the risk of heart failure. These findings highlight the critical need for further research to explore the underlying mechanisms and to inform targeted preventive measures and clinical strategies aimed at mitigating the risks associated with heavy metal exposure in vulnerable populations.

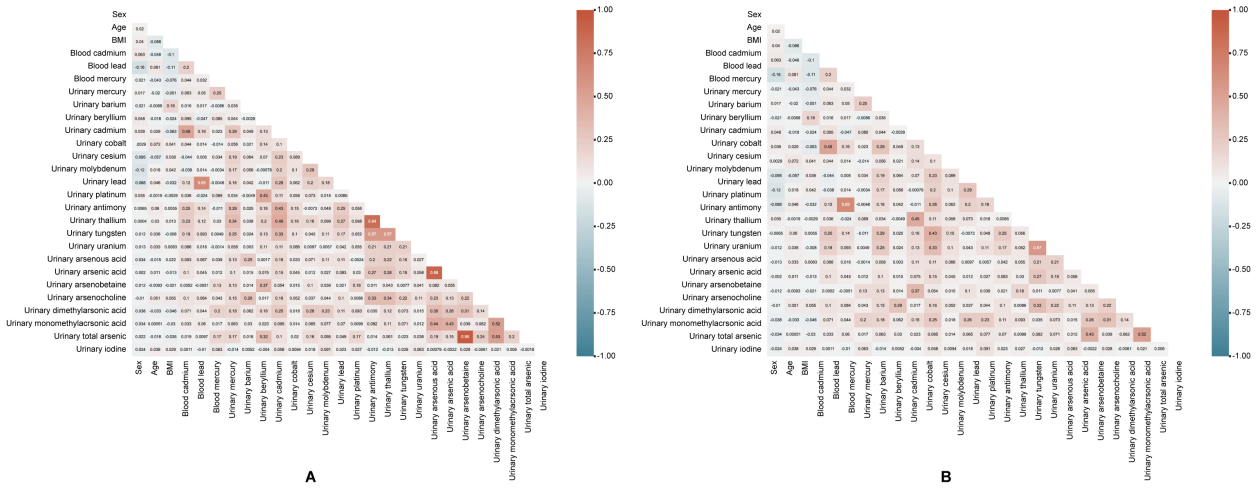
References

1. Järup L. Hazards of heavy metal contamination. Br Med Bull. 2003;68:167-82.
2. Lamas GA, Bhatnagar A, Jones MR, Mann KK, Nasir K, Tellez-Plaza M, et al. Contaminant Metals as Cardiovascular Risk Factors: A Scientific Statement From the American Heart Association. J Am Heart Assoc. 2023;12:e029852.
3. Lind L, Araujo JA, Barchowsky A, Belcher S, Berridge BR, Chiamvimonvat N, et al. Key Characteristics of Cardiovascular Toxicants. Environ Health Perspect. 2021;129:95001.
4. Singh P, O'Toole TE, Conklin DJ, Hill BG, Haberzettl P. Endothelial progenitor cells as critical mediators of environmental air pollution-induced cardiovascular toxicity. Am J Physiol Heart Circ Physiol. 2021;320:H1440-H55.
5. Pan Z, Gong T, Liang P. Heavy Metal Exposure and Cardiovascular Disease. Circ Res. 2024;134:1160-78.

6. Jiang Q, Xiao Y, Long P, Li W, Yu Y, Liu Y, et al. Associations of plasma metal concentrations with incident dyslipidemia: Prospective findings from the Dongfeng-Tongji cohort. Chemosphere. 2021;285:131497.
7. Marques M, Millás I, Jiménez A, García-Colis E, Rodriguez-Feo JA, Velasco S, et al. Alteration of the soluble guanylate cyclase system in the vascular wall of lead-induced hypertension in rats. J Am Soc Nephrol. 2001;12:2594-600.
8. Fu Z, Xi S. The effects of heavy metals on human metabolism. Toxicol Mech Methods. 2020;30:167-76.
9. Zhang Y, Huo X, Lu X, Zeng Z, Faas MM, Xu X. Exposure to multiple heavy metals associate with aberrant immune homeostasis and inflammatory activation in preschool children. Chemosphere. 2020;257:127257.
10. Zheng K, Zeng Z, Tian Q, Huang J, Zhong Q, Huo X. Epidemiological evidence for the effect of environmental heavy metal exposure on the immune system in children. Sci Total Environ. 2023;868:161691.
11. Riffo-Campos AL, Fuentes-Trillo A, Tang WY, Soriano Z, De Marco G, Rentero-Garrido P, et al. In silico epigenetics of metal exposure and subclinical atherosclerosis in middle aged men: pilot results from the Aragon Workers Health Study. Philos Trans R Soc Lond B Biol Sci. 2018;373.
12. Wang C, Pi X, Yin S, Liu M, Tian T, Jin L, et al. Maternal exposure to heavy metals and risk for severe congenital heart defects in offspring. Environ Res. 2022;212:113432.
13. Sun J, Mao B, Wu Z, Jiao X, Wang Y, Lu Y, et al. Relationship between maternal exposure to heavy metal titanium and offspring congenital heart defects in Lanzhou, China: A nested case-control study. Front Public Health. 2022;10:946439.
14. Fu Y, Liu Y, Liu Y, Wang Y, Zhu M, Lin W, et al. Relationship between cumulative exposure to metal mixtures and heart rate among Chinese preschoolers. Chemosphere. 2022;300:134548.
15. Wang X, Han X, Guo S, Ma Y, Zhang Y. Associations between patterns of blood heavy metal exposure and health outcomes: insights from NHANES 2011-2016. BMC Public Health. 2024;24:558.
16. Liu Q, Xu C, Jin J, Li W, Liang J, Zhou S, et al. Early-life exposure to lead changes cardiac development and compromises long-term cardiac function. Sci Total Environ. 2023;904:166667.
17. Sears CG, Eliot M, Raaschou-Nielsen O, Poulsen AH, Harrington JM, Howe CJ, et al. Urinary Cadmium and Incident Heart Failure: A Case-Cohort Analysis Among Never-Smokers in Denmark. Epidemiology. 2022;33:185-92.
18. Tomasoni D, Adamo M, Lombardi CM, Metra M. Highlights in heart failure. ESC Heart Fail. 2019;6:1105-27.
19. Savarese G, Becher PM, Lund LH, Seferovic P, Rosano GMC, Coats AJS. Global burden of heart failure: a comprehensive and updated review of epidemiology. Cardiovasc Res. 2023;118:3272-87.
20. Bozkurt B, Ahmad T, Alexander KM, Baker WL, Bosak K, Breathett K, et al. Heart Failure Epidemiology and Outcomes Statistics: A Report of the Heart Failure Society of America. J Card Fail. 2023;29:1412-51.
21. Peters JL, Perlstein TS, Perry MJ, McNeely E, Weuve J. Cadmium exposure in association with history of stroke and heart failure. Environ Res. 2010;110:199-206.
22. Tellez-Plaza M, Jones MR, Dominguez-Lucas A, Guallar E, Navas-Acien A. Cadmium exposure and clinical cardiovascular disease: a systematic review. Curr Atheroscler Rep. 2013;15:356.
23. Tellez-Plaza M, Guallar E, Howard BV, Umans JG, Francesconi KA, Goessler W, et al. Cadmium exposure and incident cardiovascular disease. Epidemiology. 2013;24:421-9.
24. Deering KE, Callan AC, Prince RL, Lim WH, Thompson PL, Lewis JR, et al. Low-level cadmium exposure and cardiovascular outcomes in elderly Australian women: A cohort study. Int J Hyg Environ Health. 2018;221:347-54.
25. Guo J, Su L, Zhao X, Xu Z, Chen G. Relationships between urinary antimony levels and both mortalities and prevalence of cancers and heart diseases in general US population, NHANES 1999-2010. Sci Total Environ. 2016;571:452-60.
26. Packer M. Cobalt Cardiomyopathy: A Critical Reappraisal in Light of a Recent Resurgence. Circ Heart Fail. 2016;9.
27. Li X, Zhao Y, Zhang D, Kuang L, Huang H, Chen W, et al. Development of an interpretable machine learning model associated with heavy metals' exposure to identify coronary heart disease among US adults via SHAP: Findings of the US NHANES from 2003 to 2018. Chemosphere. 2023;311:137039.
28. Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. Journal of Internal Medicine. 2018;284:603-19.
29. Vallabhaneni R, Dontu S, Pareek PK. Method for interpreting predictions of black-box machine learning models, involves performing interpretation of explanations as feature importance plots or individual instance explanations by utilizing interactive visualizations. NITTE MEENAKSHI TECHNOLOGY INST (NITT-Non-standard) VALLABHANENI R (VALL-Individual) DONTU S (DONT-Individual) PAREEK P K (PARE-Individual).
30. Van den Broeck G, Lykov A, Schleich M, Suciu D. On the Tractability of SHAP Explanations. Journal of Artificial Intelligence Research. 2022;74:851-86.
31. Mou C, Ren J. Automated ICD-10 code assignment of nonstandard diagnoses via a two-stage framework. Artif Intell Med. 2020;108:101939.
32. Stekhoven DJ, Bühlmann P. MissForest--non-parametric missing value imputation for mixed-type data. Bioinformatics. 2012;28:112-8.

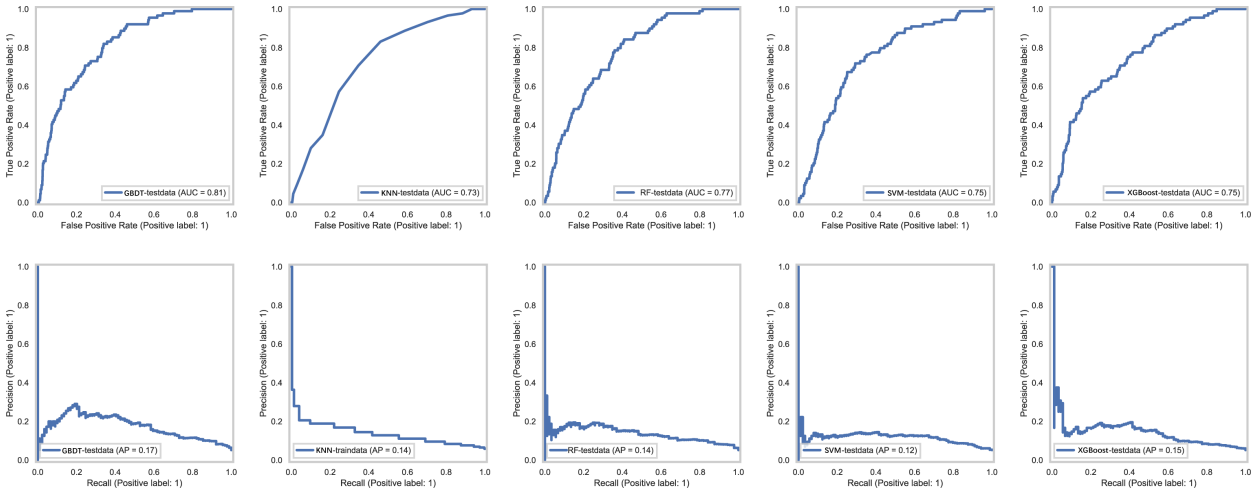
33. Chan JYL, Leow SMH, Bea KT, Cheng WK, Phoong SW, Hong ZW, et al. Mitigating the Multicollinearity Problem and Its Machine Learning Approach: A Review. Mathematics. 2022;10.
34. Gustavo E. A. P. A. Batista RCP, Maria Carolina Monard. A study of the behavior of several methods for balancing machine learning training data. SIGKDD Explor Newsl. 2004;6:20 – 9.
35. Lui TCC, Gregory DD, Anderson M, Lee W-S, Cowling SA. Applying machine learning methods to predict geology using soil sample geochemistry. Applied Computing and Geosciences. 2022;16:100094.
36. Saeed MH, Hama JI. Cardiac disease prediction using AI algorithms with SelectKBest. Medical & Biological Engineering & Computing. 2023;61:3397-408.
37. Schaffernicht E, Kaltenhaeuser R, Verma SS, Gross HM. On Estimating Mutual Information for Feature Selection. ARTIFICIAL NEURAL NETWORKS-ICANN 2010, PT I2010. p. 362-+.
38. Wainer J, Cawley G. Nested cross-validation when selecting classifiers is overzealous for most practical applications. EXPERT SYSTEMS WITH APPLICATIONS. 2021;182.
39. Vij A, Nanjundan P. Comparing Strategies for Post-Hoc Explanations in Machine Learning Models. MOBILE COMPUTING AND SUSTAINABLE INFORMATICS2022. p. 585-92.
40. Friedman JH. Stochastic gradient boosting. Computational Statistics & Data Analysis. 2002;38:367-78.
41. Zhang C, Liu C, Zhang X, Almpanidis G. An up-to-date comparison of state-of-the-art classification algorithms. Expert Systems with Applications. 2017;82:128-50.
42. Wang H, Chai K, Du M, Wang S, Cai J-P, Li Y, et al. Prevalence and Incidence of Heart Failure Among Urban Patients in China: A National Population-Based Analysis. Circ Heart Fail. 2021;14:e008406.
43. Elbadawi A, Dang A, Elgendy IY, Thakker R, Albaeni A, Omer MA, et al. Age-specific trends and outcomes of hospitalizations with acute heart failure in the United States. Int J Cardiol. 2021;330.
44. Fagerberg B, Barregard L. Review of cadmium exposure and smoking-independent effects on atherosclerotic cardiovascular disease in the general population. Journal of Internal Medicine. 2021;290:1153-79.
45. Kim J, Song H, Lee J, Kim YJ, Chung HS, Yu JM, et al. Smoking and passive smoking increases mortality through mediation effect of cadmium exposure in the United States. Sci Rep. 2023;13:3878.
46. Xing X, Xu M, Yang L, Shao C, Wang Y, Qi M, et al. Association of selenium and cadmium with heart failure and mortality based on the National Health and Nutrition Examination Survey. J Hum Nutr Diet. 2023;36:1496-506.
47. Hollins DM, McKinley MA, Williams C, Wiman A, Fillos D, Chapman PS, et al. Beryllium and lung cancer: a weight of evidence evaluation of the toxicological and epidemiological literature. Crit Rev Toxicol. 2009;39 Suppl 1.
48. Wagoner JK, Infante PF, Bayliss DL. Beryllium: an etiologic agent in the induction of lung cancer, nonneoplastic respiratory disease, and heart disease among industrially exposed workers. Environ Res. 1980;21:15-34.
49. Ward E, Okun A, Ruder A, Fingerhut M, Steenland K. A mortality study of workers at seven beryllium processing plants. Am J Ind Med. 1992;22:885-904.
50. Cheung AC, Banerjee S, Cherian JJ, Wong F, Butany J, Gilbert C, et al. Systemic cobalt toxicity from total hip arthroplasties: review of a rare condition Part 1 - history, mechanism, measurements, and pathophysiology. Bone Joint J. 2016;98-B.
51. Wyles CC, Wright TC, Bois MC, Amin MS, Fayyaz A, Jenkins SM, et al. Myocardial Cobalt Levels Are Elevated in the Setting of Total Hip Arthroplasty. J Bone Joint Surg Am. 2017;99:e118.
52. Nigra AE, Howard BV, Umans JG, Best L, Francesconi KA, Goessler W, et al. Urinary tungsten and incident cardiovascular disease in the Strong Heart Study: An interaction with urinary molybdenum. Environ Res. 2018;166:444-51.
53. Shiue I. Urinary environmental chemical concentrations and vitamin D are associated with vision, hearing, and balance disorders in the elderly. Environ Int. 2013;53:41-6.
54. Oyagbemi AA, Omobowale TO, Asenuga ER, Abiola JO, Adedapo AA, Yakubu MA. Kolaviron attenuated arsenic acid induced-cardiorenal dysfunction via regulation of ROS, C-reactive proteins (CRP), cardiac troponin I (CTnI) and BCL2. J Tradit Complement Med. 2018;8:396-409.
55. Yu Y-J, Li Z-C, Tian J-L, Hao C-J, Kuang H-X, Dong C-Y, et al. Why Do People Gain Belly Fat in Rural Areas? A Study of Urinary Metal(loid)s and Abdominal Obesity in China. Environ Sci Technol. 2023;57:7938-49.
56. Shen T, Zhong L, Ji G, Chen B, Liao M, Li L, et al. Associations between metal(loid) exposure with overweight and obesity and abdominal obesity in the general population: A cross-sectional study in China. Chemosphere. 2024;350:140963.
57. Wang X, Mukherjee B, Park SK. Associations of cumulative exposure to heavy metal mixtures with obesity and its comorbidities among U.S. adults in NHANES 2003-2014. Environ Int. 2018;121:683-94.
58. Wang M, Xu Y, Pan S, Zhang J, Zhong A, Song H, et al. Long-term heavy metal pollution and mortality in a Chinese population: an ecologic study. Biol Trace Elem Res. 2011;142:362-79.
59. Liu Y, Jin Z, Fu S. Threshold and combined effects of heavy metals on the risk of phenotypic age acceleration among U.S. adults. Biometals. 2024.

Supplementary Materials

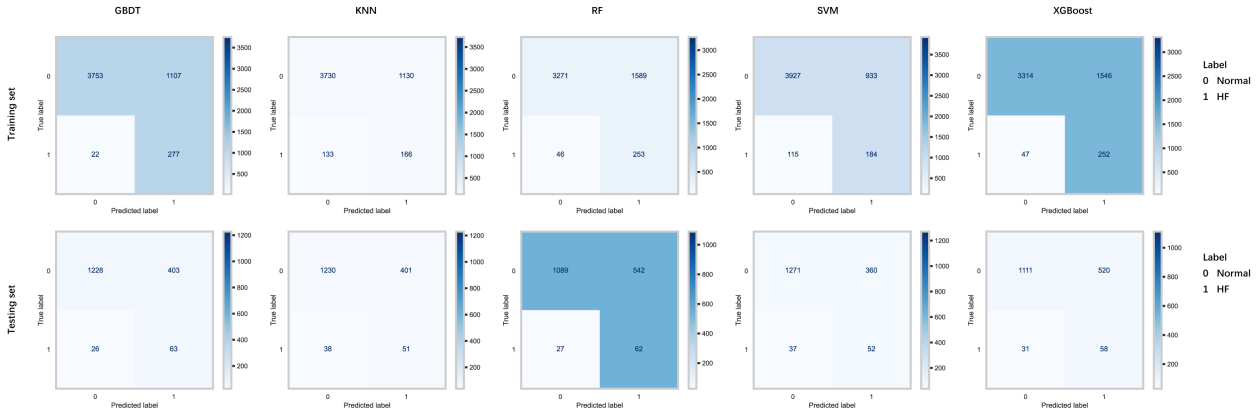


Supplementary Figure 1 | Correlation Heatmap of Features

This figure displays a correlation heatmap for each feature, with the numbers in each grid representing the Pearson correlation coefficients. Values close to +1 or -1 indicate strong positive or negative correlations, respectively, while values near 0 suggest little to no correlation. Panel A shows the heatmap prior to the removal of highly correlated variables, whereas Panel B presents the heatmap after these variables have been eliminated.



Supplementary Figure 2 | ROC and PR curves of the five models in the testing set



Supplementary Figure 3 | Confusion matrix of the five models

Each cell in the matrix indicates the number of instances classified into each category.