



Tiansheng Xia¹ and Kaiyu Han^{1*}

Abstract

Background Chronic bronchitis (CB), as a core precursor of Chronic Obstructive Pulmonary Disease (COPD), is crucial for global disease burden prevention and control. Although the association between heavy metal exposure and respiratory damage has been preliminarily demonstrated, traditional linear models are difficult to resolve the nonlinear interactions and dose–response heterogeneity. The aim of this study was to construct the first heavy metal exposure-chronic bronchitis risk prediction model by integrating exposureomics data through machine learning (ML).

Methods Weighted logistic regression was used to assess the association of 14 blood and urine heavy metals with CB based on nationally representative samples from the 2005–2015 National Health and Nutrition Examination Survey (NHANES). The Boruta algorithm was further applied to screen the characteristic variables and construct 10 ML models. The best model was selected by four evaluation metrics: accuracy, specificity, sensitivity, and area under the ROC curve (AUC), and the best model was visually interpreted using Shapley's additive interpretation (SHAP).

Results The multifactorial logistic regression model showed that urinary cadmium (OR = 1.53, 95% CI = 1.17–1.98) versus blood cadmium (OR = 1.36, 1.13-1.65) was an independent risk factor for CB. The CatBoost model had the best predictive performance (AUC = 0.805), with smoking as the most significant predictor, followed by blood cadmium concentration and gender.

Conclusion In this research, the first risk prediction diagnostic model for heavy metal-chronic bronchitis was developed, in which CatBoost model had the best performance, and it provides a referenceable prediction model for the screening of high-risk groups.

Keywords Machine Learning, Heavy Metal Exposure, Chronic Bronchitis, NHANES

Introduction

CB is a respiratory disease characterised by persistent cough and sputum as the main clinical symptoms, which are diagnosed by cough and sputum symptoms lasting at least 3 months per year and occurring for 2 or more consecutive years. Notably, despite the clinically significant association between CB and chronic obstructive pulmonary disease (COPD), some patients with CB still

*Correspondence: Kaiyu Han

hankaiyu2002@163.com

¹ Department of Respiratory and Critical Care Medicine, The Second Affiliated Hospital of Harbin Medical University, 246 Xuefu Road, Nangang District, Harbin 150001, China



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

have lung function in the normal range. The main risk factors for CB include chronic smoking, air pollution, and occupational exposure to hazardous gases, particles, or biomass fuel [1-3]. Epidemiological evidence further suggests that chronic respiratory diseases (CRDs) globally totalled 77.6 million cases in 2019, resulting in approximately 4 million deaths, with COPD dominating the picture [4]. As a core precursor phenotype of COPD progression, early identification and intervention of CB is strategically important to reduce the global disease burden.

Heavy metal exposure, as a harmful environmental pollutant, is widely present in water, soil, air, and food, and has become a serious public health problem worldwide. Heavy metal exposure is closely linked to many diseases and aggravates the risk of predisposition to various diseases, such as cardiovascular diseases [5], immune system diseases [6], and also neurotoxicity [7], and hepatotoxicity [8]. Heavy metal exposure is also closely linked to the respiratory system [9–11]. Heavy metal exposure leads to apoptosis and tissue damage in the lungs through enhanced oxidative stress and inflammatory response, which in turn destroys alveolar structure, triggers chronic inflammatory response and emphysema, and ultimately increases the risk of COPD [9]. Mixed heavy metal exposure leads to lung function impairment with significant reduction in FVC and FEV1 levels [10]. Heavy metal exposure affects the genomic stability and mutation rate of lung adenocarcinoma cancer-related genes, which play an important role in the development of lung adenocarcinoma [11]. However, traditional linear regression models have difficulty in capturing the nonlinear interactions between such complex diseases and metal exposure, which severely limits the clinical applicability of existing risk assessment frameworks.

In this background, machine learning (ML) technology provides a new paradigm for solving the above methodological bottlenecks. ML, as a cutting-edge technology, has shown great potential for application in the medical field by virtue of its ability to process large-scale, complex and diverse data. The application of ML not only provides new directions for biomedical research and individualised medicine, but also brings new opportunities for computer-aided diagnosis, which is expected to significantly advance the development of global healthcare. By applying ML to clinical datasets, it is possible to develop powerful risk prediction models and optimise patient classification, thus improving the accuracy of medical decisions [12, 13]. However, the 'black-box' nature of ML algorithms makes the prediction results difficult to understand intuitively, limiting their wide application in some fields. The SHAP method, as an advanced interpretability tool, provides an intuitive and comprehensive visual interpretation by quantifying the contribution of each feature to the model prediction, thus significantly overcoming the 'black box' limitation of traditional models [14, 15].

Based on this, the study used NHANES from 2005–2015 to study the correlation between heavy metal exposure and chronic bronchitis risk, and constructed 10 machine learning models and evaluation metrics through machine learning methods to construct a diagnostic model with good performance for predicting CB risk and SHAP visualisation.

Methods

Data sources and study population

NHANES is a representative cross-sectional study conducted by the National Center for Health Statistics (NCHS) to assess the health and nutritional status of the US ambulatory population. It is conducted every two years and collects data from a variety of sources including demographic information, physical examinations, laboratory tests, and questionnaires. All participants provided written informed consent, and the study was approved by the NCHS Research Ethics Review Board [16, 17].

In this research, we collected and analysed NHANES data from 2005 to 2015, resulting in the inclusion of 7,493 participants representing the exposure-disease association profile of the US ambulatory population, with the screening and study process shown in Fig. 1. All statistical analyses integrated sampling weights (WTMEC2YR), clustering (SDMVPSU) with stratification variables (SDMVSTRA) through the R survey package to reflect national population representation.

Data collection

Urine and blood heavy metal assessment

Total of 14 metal variables were extracted in this study, including mercury (Hg), lead (Pb) and cadmium (Cd) in blood, and cobalt (Co), molybdenum (Mo), cadmium (Cd), antimony (Sb), caesium (Cs), barium (Ba), tungsten (W), thallium (Tl), lead (Pb), uranium (U) and arsenic (AS) in urine. The metal concentrations ($\mu g/L$) in whole blood and urine were measured by inductively coupled plasma mass spectrometry (ICP-MS), whereas the concentration of AS in urine was determined by inductively coupled plasma dynamic reaction cell mass spectrometry (ICP-DRC-MS). According to the standard, all metal concentration values below the detection limit (LOD) were processed by dividing by $\sqrt{2}$. Urine metal concentrations (μ g/L) were corrected for urinary creatinine (mg/dL) and expressed as micrograms per gram of creatinine ($\mu g/g$ creatinine) [18, 19]. Detailed information on the LOD and detection rates of the 14 blood or urine heavy metals is shown in



Fig. 1 Study Design Flowchart

Supplementary Table 1. To ensure the reliability of the analytical results, only those heavy metals with a detection rate > 70% were further analyzed [18]. The results showed that all 14 heavy metals included in the analysis had high detection rates (> 70%). More detailed information on laboratory procedures and data processing is available in the NHANES database.

Definition of chronic bronchitis

The definition of patients with chronic bronchitis in this study was based on two questions, (1) 'Have you ever had chronic bronchitis?' (2) 'Do you still suffer from chronic bronchitis?' [20]. Participants who answer "no" to the first question are considered to have never had chronic bronchitis. If they answer "yes" to the first and "no" to the second, they are identified as former sufferers of chronic bronchitis. A "yes" response to both questions indicates current chronic bronchitis sufferers. In this study, patients with current chronic bronchitis were used as the case group, and the rest of the participants were used as the control group.

Covariates

Multiple covariates were included in this study, including demographic variables, body mass index (BMI), smoking, alcohol consumption, hypertension, diabetes, cancer, and cardiovascular disease (CVD). Demographic variables included age, gender, race, education level, poverty-toincome ratio (PIR), and marital status. Smoking history was categorised based on two questions, 'Have you ever smoked at least 100 cigarettes in your life?' and 'Do you currently smoke?' Based on the responses, participants were categorised as non-smokers, former smokers and current smokers [21]. Drinking history was based on 'whether more than 12 bottles of alcohol were consumed in a year, and participants were categorised as drinkers or non-drinkers. Hypertension was defined as participants who had three consecutive measurements of systolic blood pressure >140 mmHg and/or diastolic blood pressure > 90 mmHg or who answered 'yes' when asked if they had been told more than twice that they had hypertension or if they were taking medication for hypertension [22]. The diagnosis of diabetes is based on three questions, (1) 'Have you been diagnosed with diabetes

by a healthcare professional?' (2) 'Are you currently using insulin?' (3) 'Are you currently taking oral hypoglycemic medication to regulate your blood sugar levels?'. CVD includes stroke, congestive heart failure, angina, and coronary heart disease [23].

Statistical analyses

Baseline analysis and logistic regression

In this research, we used the R programming environment for data processing and analysis. Continuous variables were presented in the form of mean±standard deviation, while categorical variables were expressed as percentages. For between-group comparisons, we used the Mann–Whitney U test or t test to assess differences in continuous variables, as well as survey-weighted chisquare tests and linear regression to analyse differences in categorical and continuous variables.

To assess the correlation between heavy metal exposure and the prevalence of CB, this research used univariate and multivariate logistic regression analyses and presented the results step by step through three models. Model 1 was an unadjusted model; Model 2 adjusted for demographic variables, BMI, smoking and drinking history on the basis of Model 1; and Model 3 further adjusted for potential confounders such as cancer, CVD, hypertension and diabetes on the basis of Model 2.

Variable selection

Boruta's algorithm is a supervised classification feature selection method based on Random Forest, which identifies important features that contribute significantly to the model by comparing the z-values of real features with randomly generated 'shadow features'. The algorithm forms the smallest optimal subset of features through multiple iterations and rigorous statistical tests to ensure the stability and reliability of feature selection [24]. Boruta's algorithm is suitable for high-dimensional and complex datasets, and can significantly improve the predictive ability and interpretability of the model. Therefore, in this research, we utilised the Boruta algorithm to screen variables related to the research outcome. To ensure the accuracy and reliability of the results, we set the number of trees in the random forest to 1500 to improve the accuracy of feature importance assessment. At the same time, we set the p-value threshold for judging feature significance to 0.001 in order to strictly screen out features that contribute significantly to the target variables. Finally, 25 meaningful variables were screened, and the screening results are shown in Fig. 2.

Development and evaluation of machine learning models

In order to ensure the rigour of the research and the reliability of the results, this research divides the dataset into a training set (70%) and a test set (30%). Based on the training set, we used the Boruta algorithm to screen out the key variables and constructed 10 machine learning models using these variables, specifically including logistic regression (LR), support vector machine (SVM), Gaussian plain Bayes (GNB), neural network (NN), random forest classifier (RF), k-nearest neighbours (KNN),



Fig. 2 Feature Selection for Machine Learning Using the Boruta Algorithm

adaptive augmented classifier (AdaBoost), Classification Boost (CatBoost), Light Gradient Booster Machine (LightGBM) and Extreme Gradient Boost (XgBoost). During the model development process, in order to further improve the generalisation ability of the model and the reliability of the evaluation results, the research used tenfold repeated cross-validation. Through this approach, the performance of the model on different data subsets can be assessed more comprehensively, thus providing a more reliable basis for model selection and optimisation.

Benchmarking is a key method for evaluating and comparing the performance of machine learning models, the core of which lies in evaluating multiple models through standardised datasets and adopting uniform evaluation metrics to ensure a fair comparison [25]. Based on the characteristics of the classification task, this research selected accuracy, sensitivity/recall, specificity, and AUC as the evaluation metrics for model performance. Among them, the AUC value is the most important indicator for selecting the best model, and its value ranges from 0 to 1. The larger the value, the better the predictive performance of the model.

Model Interpretation

SHAP is a game theory-based model interpretation method proposed by Lundberg and Lee in 2017 to quantify the contribution of each feature to model predictions. It evaluates the marginal contribution of features in different combinations by calculating Shapley values to enable local and global interpretation of model predictions. The SHAP method satisfies the key attributes of local accuracy, missingness, and consistency, and is able to effectively reveal the impact of features on individual predictions, while providing a visualisation of the importance of the global features to enhance the interpretability and trustworthiness of a complex model [26]. In this research, SHAP is used for the visualisation of the best model, revealing the importance of features through global interpretation and showing the contribution of different features to individual predictions.

Results

Baseline characteristics

In this research, we included 7,493 participants during 2005–2015, of which 199 were patients with chronic bronchitis. Using baseline characteristics analysis (Table 1), the research found that compared to the control group, patients with chronic bronchitis had a higher proportion of women, were older, had a higher proportion of smokers, and had a higher prevalence of diabetes, hypertension and tumours. In addition, patients with chronic bronchitis had significantly higher urinary concentrations of cadmium, cobalt, lead and uranium, and significantly higher blood concentrations of cadmium. In contrast, patients with chronic bronchitis had significantly lower urinary creatinine concentrations and significantly lower blood mercury concentrations. The differences in all of these indicators were statistically significant (p < 0.05).

Correlation between heavy metal exposure and risk of CB

In this research, we used weighted logistic regression analyses to assess the correlation between heavy metal exposure and the risk of CB. First, we performed a baseline comparison of 14 heavy metal concentrations in blood and urine (Table 1) to analyse the differences in metal exposure concentrations between the chronic bronchitis group and the non-chronic bronchitis group. The results showed that the difference in concentrations of seven heavy metals between the two groups was statistically significant (P < 0.05). Based on this, the study further conducted a multifactorial logistic regression analysis of the seven heavy metals to investigate their independent association with the risk of chronic bronchitis. The results showed that there was a significant association between some heavy metal exposures and the prevalence of chronic bronchitis, as shown in Table 2. In addition, the results of the multifactorial logistic regression analyses of blood and urine heavy metals and the risk of chronic bronchitis, which did not reach significance (P > 0.05) in the baseline analyses, can be seen specifically in Supplementary Table 2.

Without adjusting for any confounders (Model1), urinary Cd concentration (OR=2.31, 95% CI=1.33–4.02), blood Cd concentration (OR=1.95, 95% CI=1.62–2.35), and blood Pb concentration (OR=1.07, 95% CI=1.02–1.12) were all significantly correlated with the risk of chronic bronchitis (p < 0.05).

After adjusting for all potential confounders (Model3), both urinary Cd concentration (OR=1.53, 95% CI=1.17–1.98) and blood Cd concentration (OR=1.36, 95% CI=1.13–1.65) remained significantly associated with the prevalence of chronic bronchitis (P<0.05).

Specifically, the prevalence risk of CB increased significantly with increasing Cd concentrations in urine. These results suggest that CB is strongly associated with heavy metal Cd exposure, and that Cd exposure may increase the prevalence risk of CB and is one of the risk factors for CB.

Model evaluation and selection

In this research, we conducted a comprehensive evaluation of the performance of ten machine learning models, including key metrics such as accuracy, sensitivity, specificity, and AUC (Table 3). The ROC curves and their confidence intervals for 10 of these models can be seen

Table 1 Demographic Baseline Chart of NHANES Participants from 2005 to 2015

| Characteristic | N ¹ | Overall <i>N</i> = 52,434,794 ² | Non-Chronic Bronchitis N=51,114,035 ² | Chronic Bronchitis N=1,320,759 ² | <i>p</i> -valu ³ |
|-----------------------------------|----------------|--|---|--|-----------------------------|
| Gender | 7,493 | | | | 0.001 |
| Male | | 3,796 (50%) | 3,733 (51%) | 63 (32%) | |
| Female | | 3,697 (50%) | 3,561 (49%) | 136 (68%) | |
| Age | 7,493 | 46.72±(16.59) | 46.47±(16.58) | 56.40±(14.18) | < 0.001 |
| Race | 7,493 | | | | < 0.001 |
| Mexican American | | 1,153 (8.1%) | 1,142 (8.3%) | 11 (2.2%) | |
| Other Hispanic | | 726 (5.1%) | 714 (5.2%) | 12 (2.3%) | |
| Non-Hispanic White | | 3,437 (70%) | 3,306 (70%) | 131 (83%) | |
| Non-Hispanic Black | | 1,493 (10%) | 1,452 (10%) | 41 (11%) | |
| Other Race—Including Multi-Racial | | 684 (6.2%) | 680 (6.4%) | 4 (1.8%) | |
| Education Level | 7,493 | | | | 0.062 |
| Less than 9th Grade | | 766 (5.6%) | 747 (5.5%) | 19 (9.8%) | |
| 9-11th Grade | | 1,070 (10%) | 1,029 (10%) | 41 (15%) | |
| High School | | 1,679 (22%) | 1,640 (22%) | 39 (23%) | |
| Some College | | 2,208 (32%) | 2,139 (32%) | 69 (33%) | |
| College Graduate | | 1,770 (30%) | 1,739 (31%) | 31 (19%) | |
| Marital Status | 7,493 | | | | < 0.001 |
| Married | | 3,918 (55%) | 3,837 (56%) | 81 (51%) | |
| Widowed | | 542 (5.1%) | 516 (5.0%) | 26 (10%) | |
| Divorced | | 783 (10%) | 741 (10%) | 42 (17%) | |
| Separated | | 239 (2.3%) | 229 (2.2%) | 10 (2.9%) | |
| Never married | | 1,361 (18%) | 1,335 (18%) | 26 (11%) | |
| Living with partner | | 650 (8.8%) | 636 (8.8%) | 14 (7.6%) | |
| Ratio of family income to poverty | 7,493 | $3.02 \pm (1.63)$ | $3.04 \pm (1.63)$ | 2.42±(1.62) | 0.001 |
| BMI | 7,493 | 28.74±(6.50) | 28.68±(6.42) | 31.14±(8.78) | 0.034 |
| Smoke | 7,493 | | | | < 0.001 |
| Never | | 4,052 (54%) | 4,007 (54%) | 45 (27%) | |
| Former | | 1,909 (26%) | 1,845 (26%) | 64 (31%) | |
| Current smoker | | 1,532 (20%) | 1,442 (20%) | 90 (42%) | |
| Alcohol | 7,493 | | | | 0.220 |
| No | | 2,051 (21%) | 1,999 (21%) | 52 (26%) | |
| Yes | | 5,442 (79%) | 5,295 (79%) | 147 (74%) | |
| Hypertension | 7,493 | | | | < 0.001 |
| No | | 4,839 (70%) | 4,752 (70%) | 87 (45%) | |
| Yes | | 2,654 (30%) | 2,542 (30%) | 112 (55%) | |
| CVD | 7,493 | | | | < 0.001 |
| No | | 6,876 (94%) | 6,717 (94%) | 159 (83%) | |
| Yes | | 617 (6.4%) | 577 (6.1%) | 40 (17%) | |
| Diabetes | 7,493 | | | | 0.002 |
| No | | 6,408 (89%) | 6,266 (89%) | 142 (77%) | |
| Borden | | 136 (1.7%) | 129 (1.7%) | 7 (3.4%) | |
| Yes | | 949 (9.5%) | 899 (9.3%) | 50 (20%) | |
| Cancer | 7,493 | | | | < 0.001 |
| No | | 6,802 (91%) | 6,641 (91%) | 161 (81%) | |
| Yes | | 691 (9.4%) | 653 (9.2%) | 38 (19%) | |
| URXUCR | 7,493 | 115.79±(76.43) | 116.14±(76.65) | 102.36±(66.18) | 0.027 |
| URXUBA | 7,493 | 2.12±(3.36) | 2.12±(3.38) | 2.09±(2.84) | 0.386 |
| URXUCD | 7,493 | 0.31±(0.36) | $0.30 \pm (0.35)$ | $0.58 \pm (0.47)$ | < 0.001 |
| URXUCO | 7,493 | 0.51±(0.87) | 0.50±(0.87) | 0.57±(0.49) | < 0.001 |

| Characteristic | N ¹ | Overall <i>N</i> = 52,434,794 ² | Non-Chronic Bronchitis N=51,114,035 ² | Chronic Bronchitis N=1,320,759 ² | <i>p</i> -valu ³ |
|----------------|----------------|--|---|--|-----------------------------|
| URXUCS | 7,493 | 5.12±(3.51) | 5.09±(3.13) | 6.23±(10.43) | 0.399 |
| URXUMO | 7,493 | 47.76±(38.18) | 47.69±(38.21) | 50.39±(37.03) | 0.160 |
| URXUPB | 7,493 | 0.58±(0.76) | 0.58±(0.76) | 0.62 ± (0.45) | 0.006 |
| URXUSB | 7,493 | 0.07±(0.09) | 0.07±(0.09) | $0.06 \pm (0.04)$ | 0.861 |
| URXUTL | 7,493 | 0.18±(0.14) | 0.18±(0.14) | 0.17±(0.11) | 0.080 |
| URXUTU | 7,493 | 0.11±(0.21) | 0.11±(0.22) | 0.10±(0.10) | 0.173 |
| URXUUR | 7,493 | $0.01 \pm (0.04)$ | $0.01 \pm (0.04)$ | $0.01 \pm (0.01)$ | 0.020 |
| LBXBCD | 7,493 | 0.47 ± (0.53) | 0.46±(0.52) | 0.85±(0.83) | < 0.001 |
| LBXBPB | 7,493 | 1.43±(1.36) | 1.42±(1.36) | $1.60 \pm (1.15)$ | 0.020 |
| LBXTHG | 7,493 | 1.55±(2.43) | 1.57±(2.46) | 0.94±(0.97) | < 0.001 |
| URXUAS | 7,493 | 14.29±(40.55) | 14.30±(40.75) | 14.22±(32.01) | 0.131 |

Table 1 (continued)

Urine heavy metal concentrations (µg/g creatinine): URXUCR: Creatinine concentration in urine, URXUBA: Barium concentration in urine, URXUCD: Cadmium concentration in urine, URXUCO:Cobalt concentration in urine, URXUCS: Cesium concentration in urine, URXUMO: Molybdenum concentration in urine; URXUPB: Lead concentration in urine, URXUSB:Antimony concentration in urine; URXUTL: Thallium concentration in urine, URXUTU: Tungsten concentration in urine, URXUUR: Uranium concentration in urine, URXUTU: Tungsten concentration in urine, URXUUR: Uranium concentration in urine, URXUSB:Antimony concentration in urine, Blood heavy metal concentrations (µg/L):

LBXBCD: Cadmium concentration, LBXBPB: Lead concentration, LBXTHG: Total mercury concentration

¹ N not Missing (unweighted)

 2 n (unweighted) (%); Mean ± (SD)

³ Pearson's X^2: Rao & Scott adjustment; Design-based KruskalWallis test

in Fig. 3. These metrics provide an important basis for selecting the most suitable model.

In the performance evaluation of different machine learning models in the task of predicting CB, we observed significant differences in the performance of the models on the key metrics. Specifically, the KNN model demonstrated the highest specificity, however, its sensitivity was only 0.085, suggesting that the model is heavily biased towards negative class prediction and has limited practical application value. In contrast, the CatBoost model performed well and balanced across multiple key metrics, with the highest AUC value (0.805), showing excellent classification performance. The LightGBM and Logistic models had relatively high sensitivity but low specificity, which limited their application in identifying negative individuals and limited the overall performance of the models. The NeuralNetwork model has a more balanced performance in all indicators, with an AUC value at a medium level, which indicates that it has a certain advantage in classification differentiation ability. The Random-Forest model, on the other hand, performs generally in all indicators, and the overall performance of the model is poor. Meanwhile, SVM, GBM, Adaboost and other models have poor ability in identifying positive patients, with low sensitivity.

Comprehensive evaluation showed that the CatBoost model demonstrated balanced and excellent performance in terms of accuracy, sensitivity, specificity, and AUC value, and especially significantly outperformed the other models in terms of the combined ability to distinguish between positive and negative class samples (AUC=0.805). Although Logistic Regression and Light-GBM have a slight advantage in sensitivity, their overall lack of balance limits their practical application value. Therefore, CatBoost is the best model for this research.

Variable Correlation and SHAP Visualisation

Figure 4 illustrates the heat map of the correlation of the variables. The research found a strong positive correlation between smoking status and blood Cd concentration (correlation coefficient of 0.54), indicating that smoking may significantly affect blood Cd concentration. In addition, heat map analyses revealed inter-correlations between heavy metals, with some heavy metals showing more significant inter-correlations. For example, there was a significant positive correlation between urinary barium and urinary uranium (r=0.22), and a strong positive correlation between urinary thallium (r=0.31), which suggest that there may be a synergistic effect between some of the heavy metal exposures.

In this research, SHAP analysis was used to visually interpret the best model CatBoost, revealing the contribution of characteristic variables to model prediction. Specifically, the bar chart(Fig. 5 A) and the swarm chart(Fig. 5 B) emonstrated the importance of the feature variables and their mean SHAP values, with the feature variables ranked in descending order of their importance **Table 2** The multivariate logistic regression table of heavy metal exposure and chronic bronchitis

| | Model 1 OR(95%Cl) | Model 2 OR(95%CI) | Model 3 OR(95%CI) |
|--------|----------------------|----------------------|----------------------|
| URXUCD | 2.31 (1.33,4.02) | 1.52 (1.15,2.02) | 1.53 (1.17,1.98) |
| Q1 | Reference | Reference | Reference |
| Q2 | 2.07 (0.99,4.34) | 1.19 (0.54,2.63) | 1.20 (0.54,2.68) |
| Q3 | 6.33 (3.72,10.08) | 2.00 (1.09,3.67) | 2.04 (1.11,3.75) |
| URXUCO | 1.06 (0.99,1.13) | 0.97 (0.88,1.07) | 0.97 (0.87,1.07) |
| Q1 | Reference | Reference | Reference |
| Q2 | 140 (0.96,2.03) | 1.12 (0.74,1.69) | 1.09 (0.71,1.65) |
| | Model 1 OR(95%CI) | Model 2 OR(95%CI) | Model 3 OR(95%CI) |
| Q3 | 1.72 (0.70,4.23) | 1.24 (0.46,3.33) | 1.23 (0.47,3.24) |
| URXUPB | 1.05 (0.97,1.12) | 0.79 (0.54,1.18) | 0.83 (0.57,1.19) |
| Q1 | Reference | Reference | Reference |
| Q2 | 1.71 (1.03,2.82) | 1.10 (0.63,1.90) | 1.14 (0.66,1.97) |
| Q3 | 1.74 (1.15,2.66) | 0.85 (0.52,1.40) | 0.90 (0.57,1.45) |
| URXUUR | 0.90 (0.25,3.22) | 0.03 (0.00,24) | 0.04 (0.00,26.3) |
| Q1 | Reference | Reference | Reference |
| Q2 | 1.26 (0.76,2.10) | 0.99 (0.58,1.68) | 0.94 (0.56,1.60) |
| Q3 | 1.63 (1.01,2.64) | 1.06 (0.61,1.85) | 1.02 (0.58,1.80) |
| LBXBCD | 1.95 (1.62,2.35) | 1.36 (1.12,1.64) | 1.36 (1.13,1.65) |
| Q1 | Reference | Reference | Reference |
| Q2 | 1.37 (0.76,2.48) | 0.76 (0.39,1.50) | 0.79 (0.56,1.55) |
| Q3 | 3.82 (2.28,6.40) | 1.16 (0.59,2.28) | 1.19 (0.60,2.33) |
| LBXBPB | 1.07 (1.02,1.12) | 0.94 (0.81,1.10) | 0.95 (0.83,1.10) |
| Q1 | Reference | Reference | Reference |
| Q2 | 1.26 (0.79,2.01) | 0.77 (0.43,1.38) | 0.80 (0.45,1.42) |
| Q3 | 1.63 (1.05,2.54) | 0.77 (0.41,1.45) | 0.81 (0.46,1.44) |
| LBXTHG | 0.73 (0.46,1.15) | 0.80 (0.51,1.26) | 0.81 (0.53,1.25) |
| Q1 | Reference | Reference | Reference |
| Q2 | 0.53 (0.27,1.05) | 0.53 (0.27,1.04) | 0.55 (0.29,1.04) |
| Q3 | 0.45 (0.18,1.12) | 0.58 (0.22,1.53) | 0.60 (0.23,1.59) |

Urine heavy metal concentrations (µg/g creatinine):

URXUCD: Cadmium concentration in urine, URXUCO:Cobalt concentration in urine, URXUPB: Lead concentration in urine, URXUUR: Uranium concentration in urine

Blood heavy metal concentrations (µg/L):

 $\label{eq:LBXBCD: Cadmium concentration, LBXBPB: Lead concentration, LBXTHG: Total mercury concentration$

Model 1 is the baseline model with no adjustments

Model 2 adjusts for ethnicity, age, sex, PIR, body mass index, educational attainment, marriage condition, alcohol, and smoking behavior

Model 3 adds hypertension, diabetes, CVD, and cancer to the adjustments in Model 2 $\,$

and contribution. Among them, smoking status was identified as the most important feature variable with the largest mean SHAP value, followed by blood Cd concentration and gender. This indicates that smoking status, blood Cd concentration and gender contributed most significantly to the model predictions. Waterfall plots

Table 3Assessment of Performance Metrics for 10 MachineLearning Models in Predicting Chronic Bronchitis

| Machine Model | Accuracy | Sensitivity | Specificity | AUC |
|------------------|----------|-------------|-------------|-------|
| Logistic | 0.673 | 0.814 | 0.669 | 0.801 |
| SVM | 0.762 | 0.576 | 0.767 | 0.674 |
| GBM | 0.767 | 0.627 | 0.771 | 0.751 |
| NeuralNetwork | 0.701 | 0.763 | 0.7 | 0.785 |
| RandomForest | 0.686 | 0.695 | 0.686 | 0.740 |
| Xgboost | 0.773 | 0.661 | 0.777 | 0.779 |
| KNN | 0.958 | 0.085 | 0.981 | 0.533 |
| Adaboost | 0.75 | 0.542 | 0.755 | 0.649 |
| LightGBM | 0.672 | 0.763 | 0.67 | 0.719 |
| CatBoost | 0.773 | 0.695 | 0.775 | 0.805 |

(Fig. 5 C) and force diagrams (Fig. 5 D) demonstrate the visual interpretation of the contribution of the characteristic variables to the predicted values through individual cases. The waterfall plot clearly presents the ranking and level of contribution of different features to the individual prediction of CB. In this case, the yellow arrows indicate positive contribution while the brown arrows indicate negative contribution with a final cumulative predicted value of 0.0207. The force diagram presents the contribution of the feature variables in a different way, where the colour and length of the arrows represent the positive or negative contribution of the feature variables to the prediction of an individual and the magnitude of their contribution, respectively, which ultimately results in the predicted output value of the individual.

SHAP dependency plots further revealed the three most important variables for predicting Chronic Bronchitis: smoking, blood Cd concentration, and gender (Fig. 6 A-C). The results showed that smoking, increased blood Cd concentration, and females were all associated with an increased risk of CB, suggesting that these factors are potential risk factors for CB.

Discussion

In this research, we found that heavy metal exposure is correlated with CB risk and constructed a machine learning model with good predictive performance by integrating nationally representative exposure group data with interpretable machine learning. The study found that patients with chronic bronchitis exhibited significantly higher cadmium exposure levels compared to non-CB individuals. Specifically, the blood and urinary cadmium concentrations in CB patients were 0.85 ± 0.83 and $0.58 \pm 0.47 \ \mu g/g$ creatinine, respectively, while the corresponding values in non-CB individuals were $0.46 \pm 0.52 \ \mu g/g$ creatinine for blood cadmium and



Fig. 3 ROC Curves for 10 Machine Learning Models on Training and Validation Sets. (A) ROC Curves for the Training Sets (B) ROC Curves for the Test Set



Fig. 4 Heatmap of Variable Correlations



Fig. 5 SHAP Analysis of the Best Model (A) Bar Plot for Global Feature Importance (B) Beeswarm Plot for Feature Contributions Across Samples (C) Waterfall Plot for Feature Contributions for a Single Instance (D) Force Plot for Compact Feature Contributions for a Single Instance



Fig. 6 SHAP Dependence Plots for the Top 3 Most Important Features.(**A**) Smoking Status (**B**) Blood Cadmium Concentration (μg/L) (**C**) Gender. Note: Gender was coded as 0 for males and 1 for females; smoking status was coded as 0 for non-smokers, 1 for former smokers, and 2 for current smokers

 $0.30 \pm 0.35 \ \mu$ g/g creatinine for urinary cadmium. Further analysis showed that urinary cadmium (OR=1.53 [1.17–1.98]) and blood cadmium (OR=1.36 [1.13–1.65]) were independent risk factors for CB. In addition, the CatBoost model (AUC=0.805) achieved accurate prediction of heavy metal-chronic bronchitis risk, and its SHAP visualisation showed that blood cadmium, smoking

behaviour, and gender were significant variables predicting the risk of chronic bronchitis, which may provide a valuable combination of biomarkers for clinical screening.

In this research, we found that smoking and gender are important risk factors for chronic bronchitis. Therefore, we performed a stratified analysis of the association of smoking and gender with chronic bronchitis, the results of which are presented in Supplementary Table 3. The stratified analysis showed that women had a higher risk of developing chronic bronchitis compared with male participants (OR=2.34, 95% CI=1.43-3.83). In addition, compared to non-smokers, former smokers did not have a significantly increased risk of chronic bronchitis (OR = 1.81, 95% CI = 0.94 - 3.46), whereas current smokers had a significantly increased risk of chronic bronchitis (OR = 3.68, 95% CI = 2.15 - 6.30). During the course of the research, we also assessed the relationship between other respiratory diseases, including asthma and emphysema, and heavy metal exposure (see supplementary materials, Table 4–7). The results showed that blood cadmium (OR=1.12, 95% CI: 0.92-1.35) and urinary cadmium (OR=0.88, 95% CI: 0.65-1.20) were not significantly associated with the risk of asthma, but there was a significant positive association with emphysema (blood cadmium OR=1.94, 95% CI: 1.65-2.28; urinary cadmium OR=2.18, 95% CI. 1.55-3.07). This suggests that the toxic effects of cadmium may be airway phenotype specific, there are differences in effects on different respiratory diseases.

In recent years, high concentrations of heavy metals in the atmosphere, soil, hydrosphere and biosphere have become a global problem [27]. Heavy metal exposure has an increasing impact on human health and may lead to severe damage to the nervous system, kidneys, and metabolic system [7, 28] and increase the risk of chronic diseases [29, 30]. Therefore, accurate assessment of heavy metal exposure levels in humans is essential for public health. Commonly used assays include measurement of heavy metal concentrations by blood and urine samples. Whole blood is a good biomarker for mercury, cadmium, lead, and arsenic, but is not suitable for barium and uranium [31]. Cadmium has a biological half-life of about 2 to 3 months [32], and cadmium concentrations in whole blood are usually reflective of recent exposure levels, especially in the case of occupational exposures, where concentrations can increase rapidly [33]. Accumulation of cadmium in the body affects blood cadmium concentrations, but blood cadmium levels do not decrease to pre-exposure levels after exposure has ceased. Blood cadmium concentrations remain a valid assessment tool for some time after exposure is terminated [33]. Blood lead concentration is a reliable indicator for assessing recent lead exposure [34]. In addition, whole blood can reflect long-term molybdenum intake and chronic thallium exposure [35, 36]. Urine is the most commonly used biomarker of metal exposure and is suitable for assessing exposure levels of antimony, uranium and arsenic [31], and also reflects short-term thallium, barium and molybdenum exposure [35, 37, 38]. Due to the advantages of its noninvasive operation, urine testing can be used as a long-term biomonitoring tool for lead levels [34]. In addition, emerging biomarkers such as toenails, teeth, and hair can be used to assess long-term exposure to essential and non-essential metal levels, providing a reliable basis for heavy metal exposure [31].

Heavy metal cadmium exposure mainly comes from occupational exposure and environmental exposure. In occupational exposure, cadmium mainly exists in the form of fumes or dust, and its main exposure route is the respiratory system, which is commonly found in industrial places such as smelters, pigment factories and battery factories [33]. Environmental exposure, on the other hand, includes smoking, diet, air, and drinking water. Among them, smoking is an important source of cadmium exposure, with each cigarette containing about 1-2 pg of cadmium; smokers have 4-5 times higher cadmium concentrations in their bodies compared to nonsmokers [39–41]. In non-smokers, diet is usually the main source of cadmium exposure, and cadmium is widely present in all types of food [42]. Industrial emissions and sewage sludge contamination lead to elevated levels of cadmium in soils, which in turn increases cadmium accumulation in crops and vegetables [33]. Dietary cadmium intake is higher in women who are non-smokers and who consume mainly cereals, root vegetables or shellfish compared to women who consume a mixed diet [43, 44].

Cadmium, as a toxic heavy metal, enters the human body mainly through air, water, soil and food, and causes serious health damage through long-term accumulation in the kidneys, liver and bones [45]. In this study, cadmium exposure was found to be one of the risk factors for the risk of developing chronic bronchitis, a finding consistent with the conclusions of previous research by Humairat H. Rahman [46]. Cadmium exposure induces the accumulation of intracellular reactive oxygen species (ROS), which in turn activates mitogen-activated protein kinase (MAPK) signalling pathways, including c-Jun N-terminal kinase (JNK), extracellular signalregulated kinase (ERK), and p38 pathways. Activation of these pathways further triggers the mitochondrial apoptotic pathway, which is manifested by activation of caspase-9 and imbalance of Bcl-2/Bax ratio, ultimately leading to apoptosis of bronchial epithelial cells [47]. Meanwhile, cadmium exposure significantly up-regulates the expression of several inflammatory factors, such as interleukin-6 (IL-6), tumour necrosis factor- α (TNF- α) and interleukin-1 β (IL-1 β), which promotes cellular infiltration of lung tissues and exacerbation of inflammatory responses [48, 49]. In addition, cadmium exposure promotes peribronchial fibrosis and lung remodelling by inducing wave protein phosphorylation, activating the SMAD signalling pathway and up-regulating the expression of a variety of inflammation- and fibrosis-related mediators [50].

The synergistic effects of cadmium exposure and smoking may impair lung and airway health through multiple mechanisms. Cadmium exposure may exacerbate smoking-associated lung disease by inhibiting phagocytosis in macrophages and causing an imbalance in oxidative stress [51]. In addition, low-dose Cd exposure significantly upregulates ANO1 (Anoctamin 1) expression in airway epithelial cells by downregulating miR-381 expression, which in turn affects functions such as mucus secretion and fibroblast differentiation [52]. Cadmium also induces endoplasmic reticulum stress and inflammatory responses in bronchial epithelial cells through activation of the CCAAT enhancer-binding protein (C/ EBP) signalling pathway and its downstream target gene, DDIT3, which promotes smoking-related lung diseases [53]. These mechanisms may explain the finding of high risk of cadmium-smoking co-exposure in the research.

Machine learning has demonstrated unique advantages in processing complex, high-dimensional medical data, advancing global healthcare and making significant contributions to early health management and disease prevention. Different of these machine learning models have their own unique advantages when dealing with different types of data. Logistic regression model is a traditional method used for clinical predictive modelling with the advantages of simplicity, transparency, and interpretability [54], KNN enables flexible classification fitting; SVMs have higher memory efficiency and lower computational cost [55]. RF and XGBoost are suitable for large datasets, LightGBM is fast and has a low memory footprint. Cat-Boost excels in handling categorical variables and preventing overfitting [15, 56]. Neural networks have strong nonlinear modelling capabilities, while the AdaBoost algorithm gives good classification results on generalpurpose datasets [57].

Based on this, in this research, we used machine learning techniques to process high-dimensional and complex data to deeply explore the correlation between heavy metal exposure and the risk of chronic bronchitis, and successfully constructed a high-performance Catboost prediction model and visualised it with SHAP. The model is based on a combination of blood cadmium and biomarkers such as smoking and gender, and provides an important reference for assessing and managing the risk of chronic bronchitis in people at high risk of heavy metal exposure. In addition, this research proposes a new idea of constructing a predictive diagnostic model by integrating basic demographic information, disease history, personal history, and environmental and biological pollutant data. This approach is not only applicable to the prediction of disease risk in high-risk environmentally exposed populations, but also helps to achieve early prevention of disease and health management, which is of great significance to the healthy development of public health.

Based on the significant association between cadmium exposure and chronic bronchitis risk observed in this study, as well as the characteristics of high-risk populations (females, current smokers), the following intervention strategies are recommended. At the occupational exposure level, efforts should be made to enhance cadmium level control in industrial settings, incorporate blood cadmium monitoring into occupational health surveillance, and develop predictive models to screen high-risk workers. For smoking populations, public education on the link between cadmium exposure and respiratory disease should be implemented, emphasizing the contribution of cadmium from tobacco smoke to chronic bronchitis risk, particularly among females. Additionally, promoting diversified diets can help reduce cadmium intake from food sources.

In this research, there are still some limitations in using the NHANES database to conduct research on the correlation between chronic bronchitis and heavy metal exposure. First, the NHANES database lacks detailed medical records and a gold standard for disease diagnosis, resulting in the definition of patients with chronic bronchitis relying only on questionnaires, which may introduce recall bias. Single measurements of heavy metal concentrations in blood or urine are insufficient to accurately assess long-term exposure levels. Future studies should enhance the accuracy of exposure assessment by conducting repeated measurements of heavy metal concentrations or by using biomarkers that reflect long-term exposure, such as nail or hair samples. In addition, the NHANES database is a cross-sectional design, which does not allow for the inference of causality. Although this study controlled for certain covariates, there may still be potential confounders interfering with the research results. Second, the model construction in this research was based on the NHANES database and lacked data support from other regions. Also, the model was not externally validated to verify its actual clinical efficacy in different populations. Therefore, measures such as collecting data from different regions, designing prospective studies, and designing external validation are needed to further construct more accurate and reliable diagnostic prediction models to provide a scientific basis for early diagnosis and intervention of chronic bronchitis.

Conclusion

In this research, a machine learning model with SHAP visualisation for predicting chronic bronchitis risk was constructed based on heavy metal exposure data from the NHANES. The model showed excellent predictive performance, with smoking status and blood cadmium concentration contributing the most to the model as one of the important risk factors for chronic bronchitis risk. The results of this research suggest that the incorporation of environmental pollutants into diagnostic models can help to achieve early management and prevention of the disease.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12890-025-03724-8.

Additional file 1: Supplementary Table 1: Detection rate and limit of detection for blood and urinary heavy metals in our study. Detection rate and limit of detection for blood and urinary heavy metals in our study. This table presents the detection rates and limits of detection (LOD) for blood and urinary heavy metals analyzed in our study. Blood heavy metal concentrations are expressed in $\mu g/L$, while urinary heavy metal concentrations are expressed in $\mu g/L$, while urinary heavy metal concentrations are expressed in $\mu g/L$, while urinary heavy metal concentrations are expressed in $\mu g/Q$ creatinine. The table includes data for mercury (Hg), lead (Pb), cadmium (Cd) in blood, and cobalt (Co), molybdenum (Mo), cadmium (Cd), antimony (Sb), cesium (Cs), barium (Ba), tungsten (W), thallium (TI), lead (Pb), uranium (U), and arsenic (As) in urine.

Additional file 2: Supplementary Table 2: Multivariate Logistic Regression Analysis of Heavy Metal Exposure (with Baseline *P*>0.05) and Chronic Bronchitis. Multivariate Logistic Regression Analysis of Heavy Metal Exposure (with Baseline *P*>0.05) and Chronic Bronchitis. This table presents the results of multivariate logistic regression analyses examining the association between heavy metal exposure (*P*>0.05) and chronic bronchitis in NHANES participants. The analysis includes various heavy metals measured in urine, expressed as µg/g creatinine. Three models are presented: Model 1 is the unadjusted baseline model; Model 2 adjusts for demographic factors (ethnicity, age, sex), socioeconomic status (PIR, educational attainment, marriage condition), and lifestyle factors (alcohol and smoking behavior); Model 3 further adjusts for comorbidities (hypertension, diabetes, CVD, and cancer).

Additional file 3: Supplementary Table 3: Stratified Analysis of Chronic Bronchitis by Gender and Smoking Status. Stratified Analysis of Chronic Bronchitis by Gender and Smoking Status. This table provides a stratified analysis of chronic bronchitis risk by gender and smoking status among NHANES participants. The analysis includes odds ratios (OR) and 95% confidence intervals (Cl) for each category of smoking behavior and gender. The results highlight the differential risk of chronic bronchitis associated with smoking status and gender

Additional file 4: Supplementary Table 4: Demographic Baseline Chart of NHANES Participants from 2005 to 2015. Demographic Baseline Chart of NHANES Participants from 2005 to 2015. This table provides detailed information of NHANES participants included in the study, including demographic data, examination data, disease data, and heavy metal exposure data. Urine heavy metal concentrations are presented as µg/g creatinine, while blood heavy metal concentrations are presented as µg/L. This table focuses on asthma and non-asthma participants.

Additional file 5: Supplementary Table 5: The multivariate logistic regression table of Chromium exposure and asthma. The multivariate logistic regression table of Chromium exposure and asthma. The table presents the results of multivariate logistic regression analyses (Model1-Model3) assessing the association between chromium exposure and asthma risk.

Additional file 6: Supplementary Table 6: Demographic Baseline Chart of NHANES Participants from 2005 to 2015. Demographic Baseline Chart of

NHANES Participants from 2005 to 2015. This table provides detailed information of NHANES participants included in the study, including demographic data, examination data, disease data, and heavy metal exposure data. Urine heavy metal concentrations are presented as $\mu g/g$ creatinine, while blood heavy metal concentrations are presented as $\mu g/L$. This table focuses on emphysema and non-emphysema participants.

Additional file 7: Supplementary Table 7: The multivariate logistic regression table of Heavy Metal Exposure and emphysema. The multivariate logistic regression table of Heavy Metal Exposure and emphysema. The table includes the outcomes of multivariate logistic regression analyses (Model1-Model3) evaluating the relationship between Heavy Metal Exposure and the risk of emphysema.

Acknowledgements

We sincerely thank all the participants for their valuable contributions to this study, which utilized publicly available data from prior research studies.

Submission declaration and verification

We confirm that the manuscript contains novel research findings that have not been previously published and are not presently under consideration elsewhere. All authors have thoroughly reviewed and agreed to submit this manuscript.

Clinical trial number

Not applicable.

Authors' contributions

Tiansheng Xia was in charge of the methodology design, data curation, data collection, focal analysis, visualization, and drafting the original manuscript for this study. Kaiyu Han was responsible for funding acquisition, supervision, validation, and the review and editing processes.

Funding

This work was supported by the Heilongjiang Provincial Department of Science and Technology through the Natural Science Joint Guidance Project (No. PL2024H081) and the Horizontal Cooperation Project of the Second Affiliated Hospital of Harbin Medical University (Project No. 20220815).

Data availability

This study utilized a public database from the United States, which can be accessed via the following URL:https://www.cdc.gov/nchs/nhanes/index.html.

Declarations

Ethics approval and consent to participate

This study is based on data from the National Health and Nutrition Examination Survey (NHANES), which has been approved by the Research Ethics Review Board of the National Center for Health Statistics. The study adheres to the ethical principles of the Declaration of Helsinki and its subsequent amendments. All participants have provided written informed consent. The data utilized in this work is derived from public databases, and its information is publicly accessible and permitted for unrestricted reuse under an open license.

Consent for publication

Not applicable.

Competng interests

Tiansheng Xia and Kaiyu Han declare that they have no conflict of interest.

Received: 25 February 2025 Accepted: 13 May 2025 Published online: 22 May 2025

References

- Balte PP, Chaves PHM, Couper DJ, Enright P, Jacobs DR Jr, Kalhan R, Kronmal RA, Loehr LR, London SJ, Newman AB, et al. Association of Nonobstructive Chronic Bronchitis With Respiratory Health Outcomes in Adults. JAMA Intern Med. 2020;180(5):676–86.
- Mejza F, Gnatiuc L, Buist AS, Vollmer WM, Lamprecht B, Obaseki DO, Nastalek P, Nizankowska-Mogilnicka E, Burney PGJ, collaborators B, et al. Prevalence and burden of chronic bronchitis symptoms: results from the BOLD study. Eur Respir J. 2017;50(5):1700621.
- Jarhyan P, Hutchinson A, Khaw D, Prabhakaran D, Mohan S. Prevalence of chronic obstructive pulmonary disease and chronic bronchitis in eight countries: a systematic review and meta-analysis. Bull World Health Organ. 2022;100(3):216–30.
- Collaborators GBDCRD: Global burden of chronic respiratory diseases and risk factors, 1990-2019: an update from the Global Burden of Disease Study 2019. EClinicalMedicine. 2023;59:101936.
- 5. Pan Z, Gong T, Liang P. Heavy Metal Exposure and Cardiovascular Disease. Circ Res. 2024;134(9):1160–78.
- 6. 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.
- Yu G, Wu L, Su Q, Ji X, Zhou J, Wu S, Tang Y, Li H. Neurotoxic effects of heavy metal pollutants in the environment: Focusing on epigenetic mechanisms. Environ Pollut. 2024;345:123563.
- Renu K, Chakraborty R, Myakala H, Koti R, Famurewa AC, Madhyastha H, Vellingiri B, George A, ValsalaGopalakrishnan A. Molecular mechanism of heavy metals (Lead, Chromium, Arsenic, Mercury, Nickel and Cadmium) induced hepatotoxicity - A review. Chemosphere. 2021;271:129735.
- Yan Z, Xu Y, Li K, Liu L. Heavy metal levels and flavonoid intakes are associated with chronic obstructive pulmonary disease: an NHANES analysis (2007–2010 to 2017–2018). BMC Public Health. 2023;23(1):2335.
- Wang M, Yan L, Dou S, Yang L, Zhang Y, Huang W, Li S, Lu P, Guo Y. Blood multiple heavy metals exposure and lung function in young adults: A prospective Cohort study in China. J Hazard Mater. 2023;459:132064.
- 11. Liu M, Hong Y, Duan X, Zhou Q, Chen J, Liu S, Su J, Han L, Zhang J, Niu B. Unveiling the metal mutation nexus: Exploring the genomic impacts of heavy metal exposure in lung adenocarcinoma and colorectal cancer. J Hazard Mater. 2024;461:132590.
- Handelman GS, Kok HK, Chandra RV, Razavi AH, Lee MJ, Asadi H. eDoctor: machine learning and the future of medicine. J Intern Med. 2018;284(6):603–19.
- 13. Deo RC. Machine Learning in Medicine. Circulation. 2015;132(20):1920–30.
- Tang D, Ma C, Xu Y. Interpretable machine learning model for early prediction of delirium in elderly patients following intensive care unit admission: a derivation and validation study. Front Med (Lausanne). 2024;11:1399848.
- Fu Q, Wu Y, Zhu M, Xia Y, Yu Q, Liu Z, Ma X, Yang R. Identifying cardiovascular disease risk in the US population using environmental volatile organic compounds exposure: A machine learning predictive model based on the SHAP methodology. Ecotoxicol Environ Saf. 2024;286:117210.
- Sun Y, Wang YX, Mustieles V, Shan Z, Zhang Y, Messerlian C. Blood trihalomethane concentrations and allergic sensitization: A nationwide cross-sectional study. Sci Total Environ. 2023;871:162100.
- 17. Cheng TD, Ferderber C, Kinder B, Wei YJ. Trends in Dietary Vitamin A Intake Among US Adults by Race and Ethnicity, 2003–2018. JAMA. 2023;329(12):1026–9.
- Xiao H, Liang X, Li H, Chen X, Li Y. Trends in the prevalence of osteoporosis and effects of heavy metal exposure using interpretable machine learning. Ecotoxicol Environ Saf. 2024;286:117238.
- Shen M, Zhang Y, Zhan R, Du T, Shen P, Lu X, Liu S, Guo R, Shen X. Predicting the risk of cardiovascular disease in adults exposed to heavy metals: Interpretable machine learning. Ecotoxicol Environ Saf. 2024;290:117570.
- Ganji V, Al-Obahi A, Yusuf S, Dookhy Z, Shi Z. Serum vitamin D is associated with improved lung function markers but not with prevalence of asthma, emphysema, and chronic bronchitis. Sci Rep. 2020;10(1):11542.
- Mendy A, Salo PM, Cohn RD, Wilkerson J, Zeldin DC, Thorne PS. House Dust Endotoxin Association with Chronic Bronchitis and Emphysema. Environ Health Perspect. 2018;126(3):037007.

- 22. Cai Y, Chen M, Zhai W, Wang C: Interaction between trouble sleeping and depression on hypertension in the NHANES 2005-2018. BMC Public Health. 2022;22(1):481.
- 23. Zhang Q, Xiao S, Jiao X, Shen Y. The triglyceride-glucose index is a predictor for cardiovascular and all-cause mortality in CVD patients with diabetes or pre-diabetes: evidence from NHANES 2001–2018. Cardiovasc Diabetol. 2023;22(1):279.
- Duan M, Zhao X, Li S, Miao G, Bai L, Zhang Q, Yang W, Zhao X. Metabolic score for insulin resistance (METS-IR) predicts all-cause and cardiovascular mortality in the general population: evidence from NHANES 2001–2018. Cardiovasc Diabetol. 2024;23(1):243.
- Qi X, Wang S, Fang C, Jia J, Lin L, Yuan T. Machine learning and SHAP value interpretation for predicting comorbidity of cardiovascular disease and cancer with dietary antioxidants. Redox Biol. 2025;79:103470.
- Alabi RO, Elmusrati M, Leivo I, Almangush A, Mäkitie AA. Machine learning explainability in nasopharyngeal cancer survival using LIME and SHAP. Sci Rep. 2023;13(1):8984.
- Rahman Z, Singh VP. The relative impact of toxic heavy metals (THMs) (arsenic (As), cadmium (Cd), chromium (Cr)(VI), mercury (Hg), and lead (Pb)) on the total environment: an overview. Environ Monit Assess. 2019;191(7):419.
- Rong LP, Xu YY, Jiang XY. Heavy metal poisoning and renal injury in children. Zhongguo Dang Dai Er Ke Za Zhi. 2014;16(4):325–9.
- 29. Glicklich D, Shin CT, Frishman WH. Heavy Metal Toxicity in Chronic Renal Failure and Cardiovascular Disease: Possible Role for Chelation Therapy. Cardiol Rev. 2020;28(6):312–8.
- 30. Yu S, Wang X, Zhang R, Chen R, Ma L. A review on the potential risks and mechanisms of heavy metal exposure to Chronic Obstructive Pulmonary Disease. Biochem Biophys Res Commun. 2023;684:149124.
- Martinez-Morata I, Sobel M, Tellez-Plaza M, Navas-Acien A, Howe CG, Sanchez TR. A State-of-the-Science Review on Metal Biomarkers. Curr Environ Health Rep. 2023;10(3):215–49.
- Lauwerys R, Roels H, Regniers M, Buchet JP, Bernard A, Goret A. Significance of cadmium concentration in blood and in urine in workers exposed to cadmium. Environ Res. 1979;20(2):375–91.
- Järup L, Berglund M, Elinder CG, Nordberg G, Vahter M. Health effects of cadmium exposure–a review of the literature and a risk estimate. Scand J Work Environ Health. 1998;24(Suppl 1):1–51.
- Barbosa F Jr, Tanus-Santos JE, Gerlach RF, Parsons PJ. A critical review of biomarkers used for monitoring human exposure to lead: advantages, limitations, and future needs. Environ Health Perspect. 2005;113(12):1669–74.
- Oskarsson A, Kippler M. Molybdenum a scoping review for Nordic Nutrition Recommendations 2023. Food Nutr Res. 2023;67. https://doi.org/10. 29219/fnr,v67.10326. PMID: 38187804; PMCID: PMC10770642.
- Das AK, Chakraborty R, Cervera ML, de la Guardia M. Determination of thallium in biological samples. Anal Bioanal Chem. 2006;385(4):665–70.
- Fujihara J, Nishimoto N. Thallium poisoner's poison: An overview and review of current knowledge on the toxicological effects and mechanisms. Curr Res Toxicol. 2024;6:100157.
- Harrison GE, Carr TE, Sutton A, Rundo J. Plasma concentration and excretion of calcium-47, strontium-85, barium-133 and radium-223 following successive intravenous doses to a healthy man. Nature. 1966;209(5022):526–7.
- Bensryd I, Rylander L, Högstedt B, Aprea P, Bratt I, Fåhraéus C, Holmén A, Karlsson A, Nilsson A, Svensson BL, et al. Effect of acid precipitation on retention and excretion of elements in man. Sci Total Environ. 1994;145(1–2):81–102.
- Börjesson J, Olsson M, Mattsson S. Feasibility of a fluorescent X-ray source for in vivo X-ray fluorescence measurements of kidney and liver cadmium. Ann N Y Acad Sci. 2000;904:255–8.
- Friberg L, Vahter M. Assessment of exposure to lead and cadmium through biological monitoring: results of a UNEP/WHO global study. Environ Res. 1983;30(1):95–128.
- 42. Schaefer HR, Dennis S, Fitzpatrick S. Cadmium: Mitigation strategies to reduce dietary exposure. J Food Sci. 2020;85(2):260–7.
- Berglund M, Akesson A, Nermell B, Vahter M. Intestinal absorption of dietary cadmium in women depends on body iron stores and fiber intake. Environ Health Perspect. 1994;102(12):1058–66.

- Vahter M, Berglund M, Nermell B, Akesson A. Bioavailability of cadmium from shellfish and mixed diet in women. Toxicol Appl Pharmacol. 1996;136(2):332–41.
- 45. Wang M, Chen Z, Song W, Hong D, Huang L, Li Y. A review on Cadmium Exposure in the Population and Intervention Strategies Against Cadmium Toxicity. Bull Environ Contam Toxicol. 2021;106(1):65–74.
- Rahman HH, Niemann D, Munson-McGee SH. Urinary metals, arsenic, and polycyclic aromatic hydrocarbon exposure and risk of chronic bronchitis in the US adult population. Environ Sci Pollut Res Int. 2022;29(48):73480–91.
- Cao X, Fu M, Bi R, Zheng X, Fu B, Tian S, Liu C, Li Q, Liu J. Cadmium induced BEAS-2B cells apoptosis and mitochondria damage via MAPK signaling pathway. Chemosphere. 2021;263:128346.
- Kulas J, Ninkov M, Tucovic D, Popov Aleksandrov A, Ukropina M, Cakic Milosevic M, Mutic J, Kataranovski M, Mikrov I. Subchronic Oral Cadmium Exposure Exerts both Stimulatory and Suppressive Effects on Pulmonary Inflammation/Immune Reactivity in Rats. Biomed Environ Sci. 2019;32(7):508–19.
- 49. Wang WJ, Peng K, Lu X, Zhu YY, Li Z, Qian QH, Yao YX, Fu L, Wang Y, Huang YC, et al. Long-term cadmium exposure induces chronic obstructive pulmonary disease-like lung lesions in a mouse model. Sci Total Environ. 2023;879:163073.
- 50. Skalny AV, Lima TRR, Ke T, Zhou JC, Bornhorst J, Alekseenko SI, Aaseth J, Anesti O, Sarigiannis DA, Tsatsakis A, et al. Toxic metal exposure as a possible risk factor for COVID-19 and other respiratory infectious diseases. Food Chem Toxicol. 2020;146:111809.
- Ganguly K, Levänen B, Palmberg L, Åkesson A, Lindén A. Cadmium in tobacco smokers: a neglected link to lung disease? Eur Respir Rev. 2018;27(147):170122.
- Singh P, Li FJ, Dsouza K, Stephens CT, Zheng H, Kumar A, Dransfield MT, Antony VB. Low dose cadmium exposure regulates miR-381-ANO1 interaction in airway epithelial cells. Sci Rep. 2024;14(1):246.
- Kim J, Song H, Heo HR, Kim JW, Kim HR, Hong Y, Yang SR, Han SS, Lee SJ, Kim WJ, et al. Cadmium-induced ER stress and inflammation are mediated through C/EBP-DDIT3 signaling in human bronchial epithelial cells. Exp Mol Med. 2017;49(9):e372.
- Song X, Liu X, Liu F, Wang C. Comparison of machine learning and logistic regression models in predicting acute kidney injury: A systematic review and meta-analysis. Int J Med Inform. 2021;151:104484.
- Bzdok D, Krzywinski M, Altman N. Machine learning: supervised methods. Nat Methods. 2018;15(1):5–6.
- 56 Ahn JM, Kim J, Kim K. Ensemble Machine Learning of Gradient Boosting (XGBoost, LightGBM, CatBoost) and Attention-Based CNN-LSTM for Harmful Algal Blooms Forecasting. Toxins (Basel). 2023;15(10):608.
- 57 Li K, Zhou G, Zhai J, Li F, Shao M. Improved PSO_AdaBoost Ensemble Algorithm for Imbalanced Data. Sensors (Basel). 2019;19(6):1476.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.