BMC Public Health

RESEARCH Open Access

Development of a machine learning model related to explore the association between heavy metal exposure and alveolar bone loss among US adults utilizing SHAP: a study based on NHANES 2015–2018

Jiayi Chen[1*]

Abstract

Background Alveolar bone loss (ABL) is common in modern society. Heavy metal exposure is usually considered to be a risk factor for ABL. Some studies revealed a positive trend found between urinary heavy metals and periodontitis using multiple logistic regression and Bayesian kernel machine regression. Overfitting using kernel function, long calculation period, the definition of prior distribution and lack of rank of heavy metal will affect the performance of the statistical model. Optimal model on this topic still remains controversy. This study aimed: (1) to develop an algorithm for exploring the association between heavy metal exposure and ABL; (2) filter the actual causal variables and investigate how heavy metals were associated with ABL; and (3) identify the potential risk factors for ABL.

Methods Data were collected from National Health and Nutrition Examination Survey (NHANES) between 2015 and 2018 to develop a machine learning (ML) model. Feature selection was performed using the Least Absolute Shrinkage and Selection Operator (LASSO) regression with 10-fold cross-validation. The selected data were balanced using the Synthetic Minority Oversampling Technique (SMOTE) and divided into a training set and testing set at a 3:1 ratio. Logistic Regression (LR), Support Vector Machines (SVM), Random Forest (RF), K-Nearest Neighbor (KNN), Decision Tree (DT), and XGboost were used to construct the ML model. Accuracy, Area Under the Receiver Operating Characteristic Curve (AUC), Precision, Recall, and F1 score were used to select the optimal model for further analysis. The contribution of the variables to the ML model was explained using the Shapley Additive Explanations (SHAP) method.

Results RF showed the best performance in exploring the association between heavy metal exposure and ABL, with an AUC (0.88), accuracy (0.78), precision (0.76), recall (0.83), and F1 score (0.79). Age was the most important factor in the ML model (mean| SHAP value| = 0.09), and Cd was the primary contributor. Sex had little effect on the ML model contribution.

*Correspondence:
Jiayi Chen
cjy13912736738@163.com

Full list of author information is available at the end of the article

Conclusion In this study, RF showed superior performance compared with the other five algorithms. Among the 12 heavy metals, Cd was the most important factor in the ML model. The relationship of Co & Pb and ABL are weaker than that of Cd. Among all the independent variables, age was considered the most important factor for this model. As for PIR, low-income participants present association with ABL. Mexican American and Non-Hispanic White show low association with ABL compared to Non-Hispanic Black and other races. Gender feature demonstrates a weak association with ABL. In the future, more advanced algorithms should be developed to validate these results and related parameters can be tuned to improve the accuracy of the model.

Clinical trial number not applicable.

Keywords Machine learning, Heavy metal, Alveolar bone loss, NHANES, Random forest

Introduction

The alveolar bone is a process located in the maxilla and mandible. This creates sockets to support teeth. The alveolar bone consists of the cortical bone, alveolar wall, and minute amounts of cancellous bone [1]. Due to insufficiency in the bone marrow, unlike in limbs, alveolar bone regeneration is a challenge for professionals once damage and absorption of alveolar bone occurs. Sufficient alveolar bone is the basis of normal tooth function and implantation. To solve the difficulty of inserting implants into the alveolar bone, guided bone regeneration based on autogenous graft is considered the gold standard for bone reconstruction. However, autograft graft usually prolongs the healing period and renders pain in patients [2]. Alveolar bone loss (ABL) in oral cavity can be divided into periodontitis and residue ridge absorption. Periodontitis is characterized by alveolar bone loss and soft tissue attachment loss. Residue ridge absorption of alveolar bone usually occur after tooth extraction [3]. Due to no detailed data on residue ridge absorption in National Health and Nutrition Examination Survey (NHANES), ABL refers to periodontitis in this study. ABL have given rise to a great burden on individuals' oral health and health systems. Bone is not a static organ that reshapes itself under loading conditions [4]. In addition to bite force, many factors affecting ABL include specific pathogenic bacteria [5], lipid metabolism disorder [6], hypertension, diabetes [7], and age [8]. Owing to irreversible ABL, the exploration of ABL risk and prevention of ABL deserve further investigation. Smoking is considered as an important risk for uncommunicable chronic diseases and the association of smoking and periodontitis has been proved [9]. A study has revealed that household smoking status is associated with heavy metal concentrations in children [10]. On the other hand, a multinational study demonstrated that the prevalence of alveolar bone loss was great geographical variations [11]. The distribution of heavy metals varies in different areas of soil. I try to guess whether heavy metal concentrations in blood will lead to alveolar bone loss.

Heavy metals refer to metals with density > 4.5 g/cm^3 including Cadmium, Lead, Manganese, Mercury, etc [12]. Soil is the main source of heavy metals, which can cause drinking water pollution [13]. Environmental pollution is a major challenge in modern society. Heavy metal is a pollutant causing numerous health problems, including kidney damage [14], cancer [15], bone loss [16], etc. Ma et al. reported that Cd is a widespread pollutant in nature and is toxic to enamel development, bone formation, and carcinogenicity. Cd inhibits osteoblast differentiation and promotes apoptosis of bone marrow mesenchymal stem cells through different signaling pathways [17]. The cytotoxicity of heavy metals is mediated by ER stress, apoptosis, necrosis, necroptosis, and ferroptosis [18]. A previous animal study suggested that immature rats exposed to lead under hypoxic conditions were more susceptible to ABL and fractures [19]. A cross-sectional study from South Korea National Health and Nutrition Examination Survey (KNHANES) revealed a significant association between periodontitis and serum Cd and lead concentrations [20]. According to epidemiological statistics, more than 40% adults suffered from periodontitis and the severe form of periodontitis was reported to have a prevalence of 11% globally [21]. The disease of periodontitis has put great burdens on health system and individuals. However, the relationships between heavy metals and periodontitis remain inconclusive. Few studies have focused on the nonlinear complex association between heavy metals and ABL, and which heavy metals demonstrate the most toxic effect on the alveolar bone. A previous study by Li et al. revealed that a positive trend was found between urinary heavy metals and periodontitis using multiple logistic regression and Bayesian kernel machine regression [22]. Multiple logistic regression as an old machine learning model to assess the association was characterized by limitations such as inability to evaluate the nonlinear complex association and overfitting of model when there were too many independent variables. Although their study used Bayesian kernel machine regression to optimize the results, ascending dimensionality of input data were prone to overfitting using kernel function, long calculation period, the definition of prior distribution and lack of rank of heavy metal. Optimal model on this topic still remains a controversy.

Machine learning (ML) is a subset of AI that learns from known data to predict unknown data. Unlike traditional statistical models (e.g., linear regression), ML can handle high-dimensional data and is not limited to two-dimensional data. Meanwhile, high-dimensional data can be reduced using ML method when too many independent variables in the model. For example, too many variables are included in Logistic regression and the results will be positive false due to insufficient data samples and the overfitting of model. ML method can remove irrelated variables and improve the generalization of the model. Independent variables are called features and dependent variables are called labels in the ML field. ML can be divided into supervised learning, unsupervised learning, semi-supervised learning, and reinforcement learning. According to the algorithm construction, the ML model can be categorized into single- and multi-model algorithms. Prediction, estimation, causal inference, and decision support are performed using the ML method [23]. ML is aimed at classification, regression, dimensionality reduction, and clustering of data. Owing to the increasing number of ML algorithms, ML can simulate human thinking, learning, and behavior to a certain degree. With developments in computing power, ML has attracted considerable attention from medical professionals. ML has been applied in numerous medical domains such as disease diagnosis, medical image analysis, data preprocessing, drug development, and free hands for doctors.

In this study, ML algorithms (logistic regression [LR], support vector machines [SVM], random forest [RF], decision trees [DT], K-nearest neighbor [KNN], and XGboost) were developed to explore the potential association between ABL and heavy metals combined with sociodemographic data based on the NHANES 2015–2018. Feature engineering (Least Absolute Shrinkage and Selection Operator [LASSO] regression, cross-validation, and Synthetic Minority Oversampling Technique [SMOTE]) was performed for the data processing. In order to select the optimal ML model for this topic and overcome the limitation of the method by Li et al., this study reduced the risk of overfitting and constructed six models to explore the association between ABL and heavy metals automatically using ML methods as efficiently as possible. The highlights of this study are as follows: (1) to develop an algorithm for exploring the association between heavy metal exposure and ABL; (2) filter the actual causal variables and investigate how heavy metals were associated with ABL; and (3) identify the potential risk factors for ABL.

Method

Study population

NHANES, a complex and multi-stage sampling program conducted by the Centers for Disease Control and Prevention (CDC), was the source of the data analyzed in this study. The survey was approved by the Institutional Review Board (IRB), and documented consent was obtained from the participants. Two cycles of the NHANES (2015–2016, 2017–2018) were extracted for heavy metal concentrations as well as corresponding covariables and dependent variables. Missing data for independent or dependent variables were excluded. The remaining participants were included in further analysis.

The definition of measurement of heavy metal concentrations

The heavy metals in urine (Barium [Ba] [URXUBA], Cadmium [Cd] [URXUCD], Cesium [Cs] [URXUCS], Cobalt [Co] [URXUCO], Manganese [Mn] [URXUMN], Molybdenum [Mo] [URXUMO], Lead [Pb] [URXUPB], Antimony [Sb] [URXUSB], Thallium [Tl] [URXUTL], Tin [Sn] [URXUSN], Tungsten [Tu] [URXUTU], Mercury [Hg] [URXUHG]) as continuous independent variables in the statistical model. The urine was measured by staff in a one-third subsample of participants 6 years and older. Specimens were stored under appropriate frozen (−30 °C) and then they were shipped to National Center for Environmental Health for further testing. Heavy metal concentrations were measured using inductively coupled plasma mass spectrometry after a simple dilution sample preparation step. If the analytical results were below the lower limit of detection, the heavy metal concentration was calculated as the lower limit of detection divided by the square root of the 2. More methodology details can visit NHANES official website.

The definition of alveolar bone loss

The oral health questionnaire was administered at an in-home using a computer-assisted personal interview (CAPI). The CAPI system is programmed with built-in consistency checks to reduce data entry errors and assist interviewers in defining key terms used in the questionnaire. To evaluate alveolar bone absorption, the question “Ever been told of bone loss around teeth by dentists?” [OHQ860] were asked by the participants. Participants who refused to answer or did not know to answer were excluded from the study. The answers to this question (yes/no) were divided into two groups: participants with and without ABL. The outcome variable was binary.

The definition of covariables

Several risk factors for ABL were identified by retrieving related references [24–26]. Covariable selection included sociodemographic factors, general health, and daily habits. Independent covariables included age [RIDAGEYR], sex (male/female) [RIAGENDR], marital status (yes/no) [DMDMARTL], education level (below high school, high school, or above) [DMDEDUC2], race [RIDRETH1], poverty–income ratio [INDFMPIR], BMI [BMXBMI], hypertension (yes/no) [BPQ020], diabetes mellitus (yes/no) [DIQ010], alcohol drinking (yes/no) [ALQ151], moderate recreational activities (yes/no) [PAQ665], and smoking (yes/no) [SMQ040]. The race covariables were mainly categorized into Mexican American, Non-Hispanic White, Non-Hispanic Black, and other races. Hypertension and diabetes mellitus were defined by the question "has your medical doctor told you have hypertension and diabetes mellitus" and borderline diabetes was not grouped into diabetes mellitus. People who were unknown or refused to answer whether they were told by medical doctors were excluded from the study. The answer to the question "Ever have 4/5 or more drinks every day?" classifies the participants into alcohol drink and no alcohol drink group. People who were unknown or refused to answer whether they drink every day were excluded from the study. The answer to the question "In a typical week, do you do any moderate-intensity sports, fitness, or recreational activities that cause a small increase in breathing or heart rate such as brisk walking, bicycling, swimming, or volleyball for at least 10 minutes continuously?" classifies the participants into activity and no activity group. People who were unknown or refused to answer whether they do moderate activities were excluded from the study. The question "Do you now smoke cigarettes?" was used as the cutoff for smoking and no smoking group. Participants with smoking every day and some days were categorized into smoking group and people with smoking not at all were classified into no smoking group. People who were unknown or refused to answer whether they smoke were excluded from the study. Continuous and binary covariables were included in the statistical models.

Correlation analysis and selection of heavy metals and covariables

Multicollinearity refers to the correlation between two variables or among more than two variables. For example, logistic regression as a simple ML model usually keeps other covariables unchanged and control these covariable affecting prediction outcomes. In practice, the correlation among variables occurs and it is difficult to explain dependent variable affected by one independent variable. Actually, multiple correlated independent variables jointly affect dependent variable. Hence, the occurrence of Multicollinearity will reduce the reliability of the estimated coefficients or reduce the stability and performance of the model. Owing to the binary covariables in the statistical model and non-normal distribution of heavy metal concentrations, Spearman correlation analysis was applied to detect the association among all independent variables. Due to a controversy on threshold for select variables correlated r (0.4–0.85), more restrictive threshold ($r = 0.4$) was applied to evaluate correlation in this study [27]. Another method to test multicollinearity is that the variance inflation factor, which is used to detect multicollinearity among the included features. The absence of multicollinearity was validated when the variance inflation factor was <5 to 10 [28]. Meanwhile, as too many independent variables were identified in the model, enhancing the probability of overfitting of the ML algorithm and difficulty in calculating high-dimensional data, the Least Absolute Shrinkage and Selection Operator (LASSO) regression was developed to select explanatory variables. The principle behind LASSO regression is that L1 regularization can be added to the loss function to compress the weight coefficient. The hyperparameter λ represents the regularization strength. LASSO regression with 10 fold cross-validation can identify appropriate features (independent variables are called features and dependent variables are called labels in ML).

Data preprocessing and machine learning model construction

As we all know, the imbalance of positive and negative samples will affect the performance of ML model. For example, there are 98 positive examples and 2 negative examples in a model. If all samples are judged as positive issues and the accuracy of this model can achieve 98%, this model has no any practical significance. Because of the imbalance between binary labels in initial data, SMOTE was used to improve the performance of the ML model. The standard scale of the included features was determined, and the data were randomly divided into a training set and testing set at a 3:1 ratio. Six ML algorithms were developed to predict the risk of alveolar bone absorption in the training set. The algorithms used in this study included LR, SVM, RF, KNN, DT, and XGboost. The constructed models were applied to the test set. All the aforementioned procedures were implemented in Python 3.12.0.

Model performance evaluation

The performance of the six ML models was assessed using the following indices: Accuracy, Precision, Recall, F1 score, and area under the receiver operator curve (AUC). Visualization of the model performance evaluation is presented in the Receiver Operating Characteristic (ROC) curve and confusion matrix. All calculations were performed using Python based on the Jupyter environment.

Interpretable method of constructed ML model

Unlike the coefficients in a simple linear model, the constructed ML models required explanation through "a black box". Shapley Additive ExPlanations (SHAP) is a method used to predict the ML model based on game theory. SHAP is a unified framework proposed by Lundberg and Lee [29] to interpret ML predictions, and it is a new approach to explain various black-box ML models. Development of an interpretable machine learning model associated with heavy metals exposure to diseases using SHAP has been validated in previous study [30]. Every feature in every sample was provided with a SHAP value. A positive SHAP value indicates that the features have a positive effect on the prediction probability, whereas negative SHAP values have a negative effect on the prediction. Consequently, the weighted average of the SHAP values for each feature demonstrated the overall effect of each feature on the prediction model. Macroscopic feature importance plots and macroscopic bee swarm maps were used to reveal the association between the features and prediction model.

Statistical analyses

Non-normally distributed data are expressed as median and interquartile range, whereas normally distributed data are described as mean and standard deviation. Categorical variables are described as percentages. All statistical analyses and plots were performed using R 4.3.1 and Python version 3.12.0. Statistical significance was defined as a two-tailed p value of < 0.05.

Results

Baseline characteristics of study populations

Figure 1 shows the screening flow of study participants and included participants 1057 adults without missing data. There were 224 participants with ABL and 833 participants without ABL. Table 1 presents the baseline characteristics of the study population. Males accounted for 60.93% of the study population and females for 39.07%. The age range of the patients was 30- to 80-years-old. The dominant age group is 61–70 and consists of 255 participants. Participants with hypertension comprised 47.02% of the total population, whereas 19.87% of the participants had diabetes [31, 32]. In total, 39.83% of the population reported no history of smoking. More

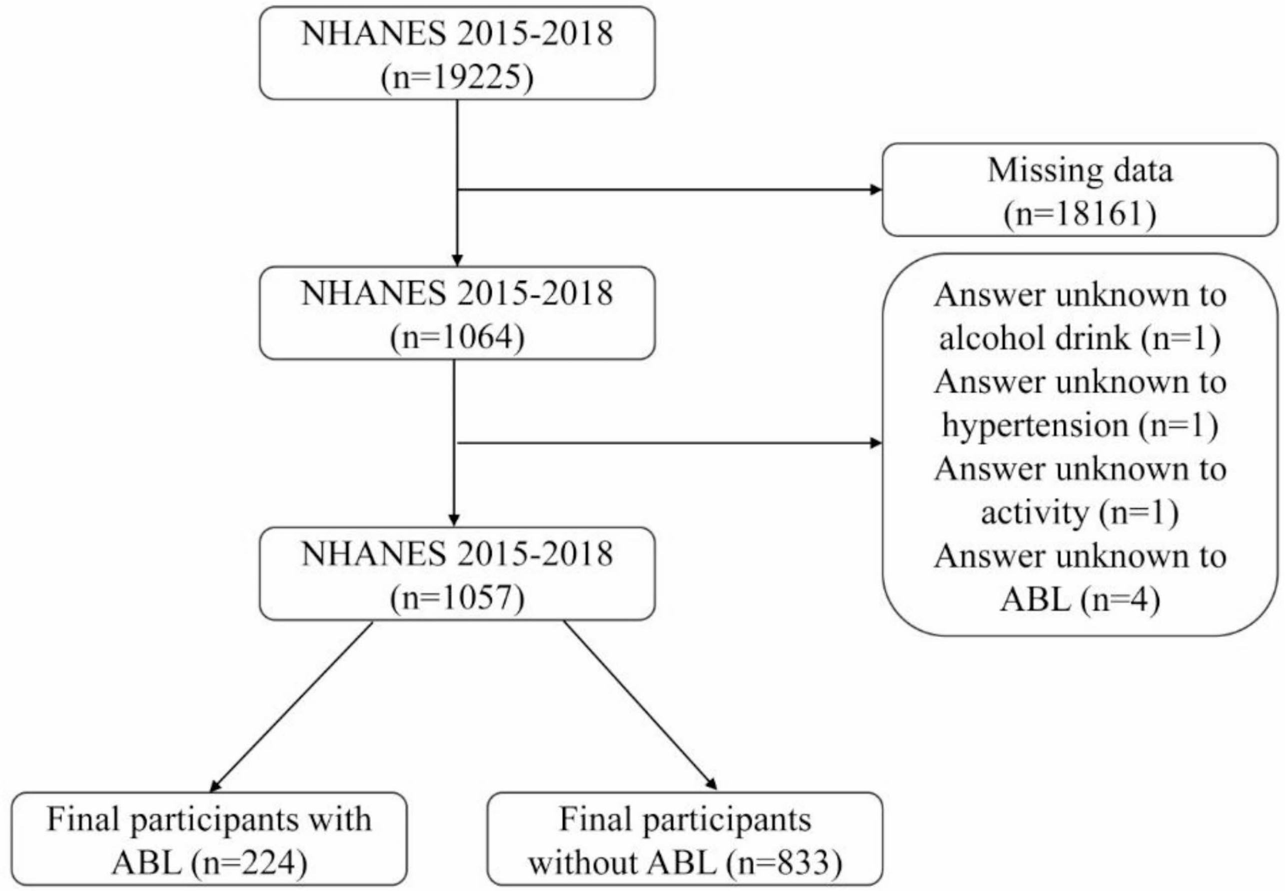
NHANES 2015-2018
(n=19225)
Missing data
(n=18161)
NHANES 2015-2018
(n=1064)
Answer unknown to alcohol drink (n=1)
Answer unknown to hypertension (n=1)
Answer unknown to activity (n=1)
Answer unknown to ABL (n=4)
NHANES 2015-2018
(n=1057)
Final participants with ABL (n=224)
Final participants without ABL (n=833)

Fig. 1 The screening flow chart of the study participants

Table 1 Baseline characteristics of the study population

		ABL	NO ABL	*p*
N		224	833	
Ba_concentration (ug/L, median [IQR])		1.17 [0.51, 2.06]	0.89 [0.44, 1.92]	0.033
Cd_concentration (ug/L, median [IQR])		0.38 [0.20, 0.67]	0.29 [0.15, 0.60]	0.002
Co_concentration (ug/L, median [IQR])		0.41 [0.26, 0.61]	0.40 [0.24, 0.61]	0.314
Cs_concentration (ug/L, median [IQR])		4.93 [3.00, 6.70]	4.33 [2.68, 6.44]	0.044
Hg_concentration (ug/L, median [IQR])		0.15 [0.06, 0.36]	0.06 [0.06, 0.32]	0.278
Mn_concentration (ug/L, median [IQR])		0.06 [0.06, 0.14]	0.06 [0.06, 0.14]	0.531
Mo_concentration (ug/L, median [IQR])		33.16 [16.73, 55.93]	34.84 [17.81, 58.77]	0.655
Pb_concentration (ug/L, median [IQR])		0.41 [0.27, 0.74]	0.39 [0.21, 0.66]	0.066
Sb_concentration (ug/L, median [IQR])		0.05 [0.03, 0.08]	0.05 [0.03, 0.08]	0.868
Sn_concentration (ug/L, median [IQR])		0.52 [0.28, 1.16]	0.50 [0.25, 1.03]	0.189
Tl_concentration (ug/L, median [IQR])		0.15 [0.10, 0.23]	0.15 [0.09, 0.23]	0.287
Tu_concentration (ug/L, median [IQR])		0.06 [0.03, 0.11]	0.05 [0.03, 0.10]	0.276
Gender (%)	Male	128 (57.1)	516 (61.9)	0.219
	Female	96 (42.9)	317 (38.1)	
Age (mean [SD])		59.60 (12.07)	55.59 (14.97)	<0.001
Race (%)	Mexican American	27 (12.1)	101 (12.1)	0.646
	Non-Hispanic White	107 (47.8)	367 (44.1)	
	Non-Hispanic Black	42 (18.8)	187 (22.4)	
	Other race	48 (21.4)	178 (21.4)	
Education (%)	Below high school	42 (18.8)	201 (24.1)	0.108
	High school or above	182 (81.2)	632 (75.9)	
Marital status (%)	Yes	136 (60.7)	502 (60.3)	0.964
	No	88 (39.3)	331 (39.7)	
PIR (mean [SD])		2.68 (1.63)	2.18 (1.47)	<0.001
Alcohol drink (%)	Yes	67 (29.9)	236 (28.3)	0.703
	No	157 (70.1)	597 (71.7)	
BMI (mean (SD))		30.31 (6.57)	29.96 (7.11)	0.504
BP (%)	Yes	115 (51.3)	382 (45.9)	0.166
	No	109 (48.7)	451 (54.1)	
DM (%)	Yes	50 (22.3)	160 (19.2)	0.346
	No	174 (77.7)	673 (80.8)	
Activity (%)	Yes	89 (39.7)	290 (34.8)	0.199
	No	135 (60.3)	543 (65.2)	
Smoking (%)	Yes	82 (36.6)	339 (40.7)	0.302
	No	142 (63.4)	494 (59.3)	

than half of the population (77.01%) had received a high school education [33]. Non-Hispanic whites were the largest race among the four races, representing 44.84% of the total participants. Less than half of the population (35.86%) preferred to engage in moderate activity during daily life.

Correlation analysis and independent variates selection

The correlation analysis between the 12 heavy metals and 12 covariables is presented in Fig. 2. As shown in Fig. 2, most heavy metals were intercorrelated. Tl and Cs ($r = 0.73$), and Tu and Mo ($r = 0.62$) showed strong positive associations though the variance inflation factor for all independent variables are between 1.05 and 1.89. From the perspective of chemistry, the heavy metal ions Cs + and Tl + are within a Van der Waal's distance of emitting trp residue in gr A in CH3OH glass at 77 K so that they are capable of inducing increased spin-orbit coupling due to a heavy atom effect [34]. Molybdenum can be incorporated into tungsten aldehyde oxidoreductase enzymes from pyrococcus furiosus [35]. This suggests multicollinearity among the independent variables. LASSO regression was performed to select appropriate variables. Figure 3 demonstrates that all the coefficients of the model tend to zero with an increase in log (λ). This indicated that the independent variables were gradually eliminated from the model. As shown in Fig. 4, all variables were excluded at lambda.1se and no coefficient was found in the model. Seven variables were included in the model for lambda.min, with the lowest prediction error in the LASSO regression model. LASSO regression with 10-fold cross-validation selected sex, race, age, PIR, Cd,

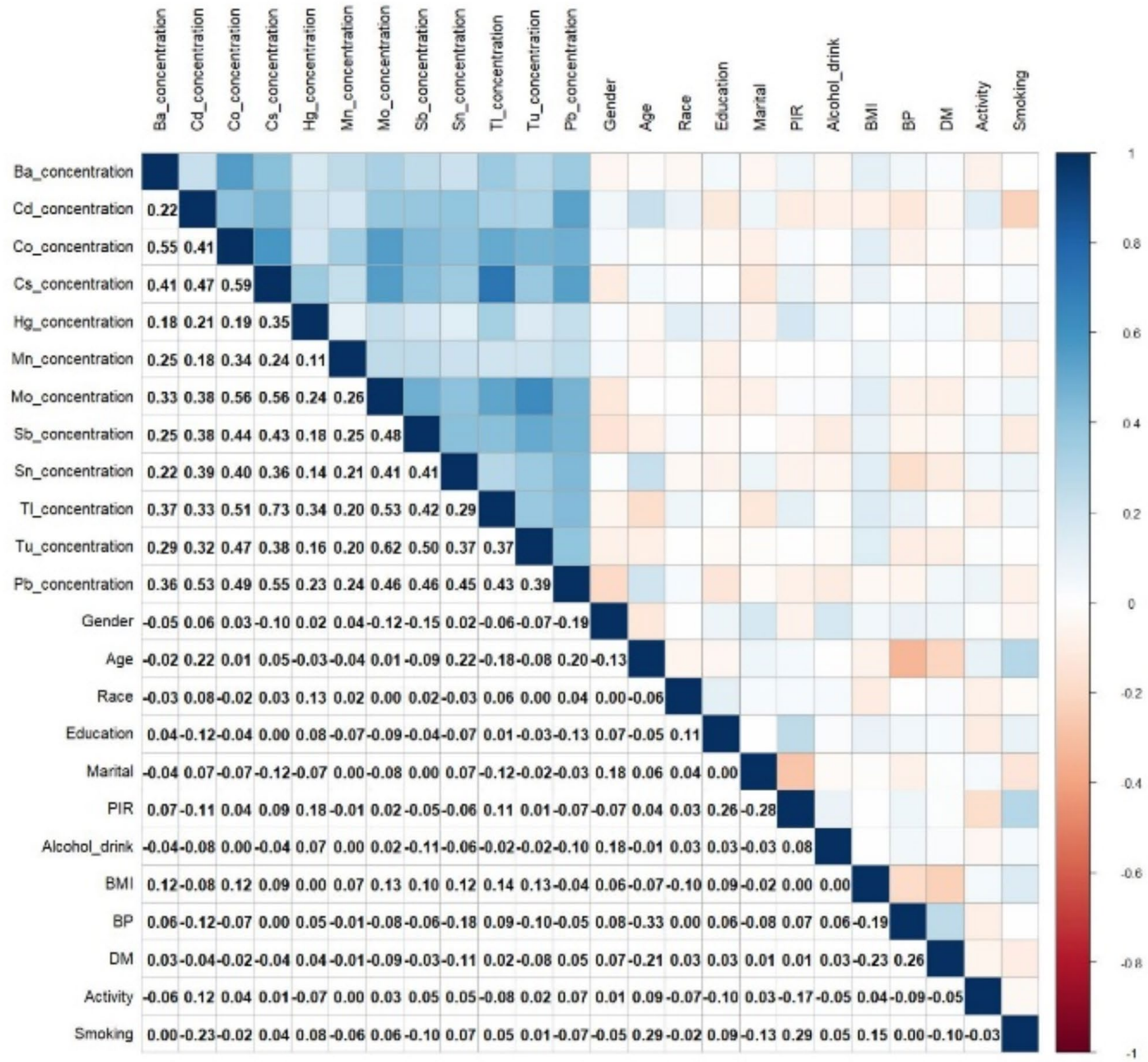

Fig. 2 The spearman correlation analysis among heavy metals and covariables

Co, and Pb as the optimal independent variables in the training set.

Evaluation of six ML models

Ultimately, 1666 participants were divided into a training set (1249) and testing set (417) after SMOTE. Figure 5 shows the ROC curves and confusion matrices for the six ML models. The ROC curve of the RF model was closest to the top-left corner, indicating its superior performance. Accuracy, AUC, Precision, Recall, and F1 score are listed in Table 2. Table 2 indicates that RF was the optimal algorithm in the training set and achieved an acceptable accuracy in the testing set. Consequently, RF was chosen for the next analysis.

Interpretable RF algorithm using SHAP

Figure 6 shows the feature importance and absolute SHAP values for each feature. Age contributed most to the ML model. Among the heavy metals, Cd is strongly associated with ABL risk. PIR showed a potential association with ABL risk. As shown in Fig. 7, most of red and bule points are concentrated to the left of the zero line in the age feature. Red points represent higher feature values and bule points represent lower feature values. This indicates that age has a negative effect on model output overall. The low values of outcome (ABL: 1, NO ABL: 2) are associated with advanced age. The PIR feature shows that most points are concentrated to the right of the zero line and PIR has a positive effect on higher

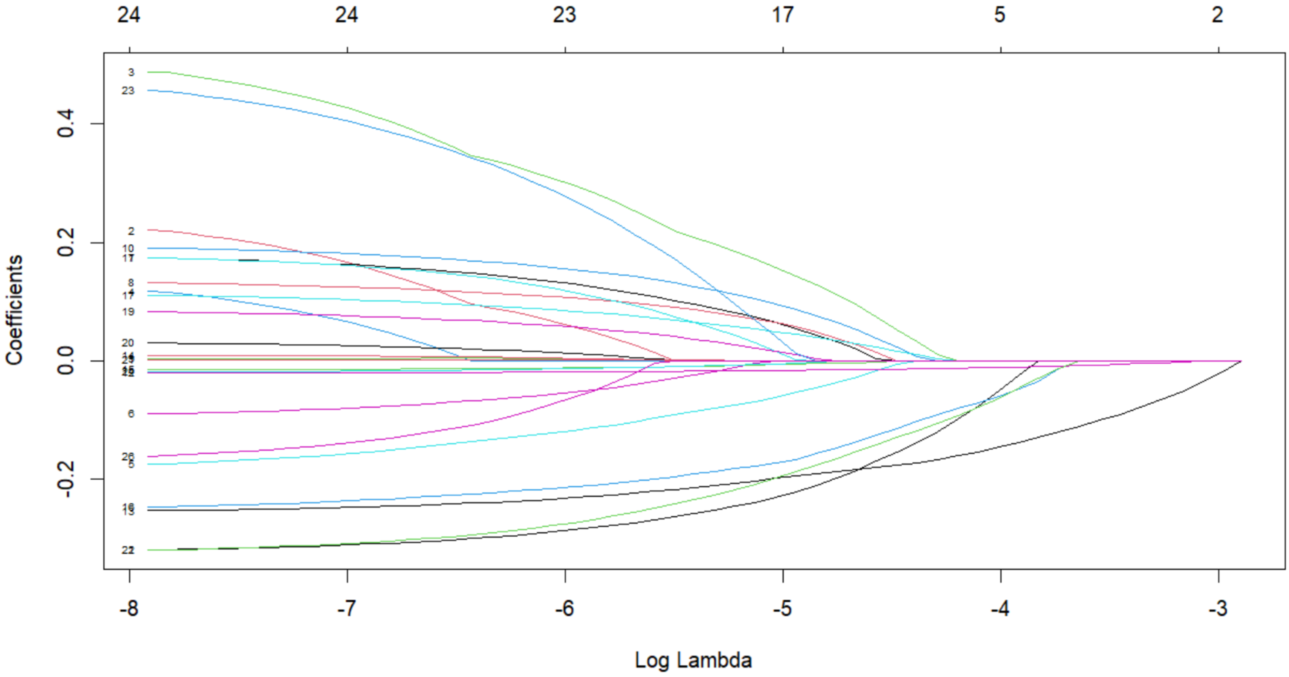

Fig. 3 The LASSO regression for variate selection

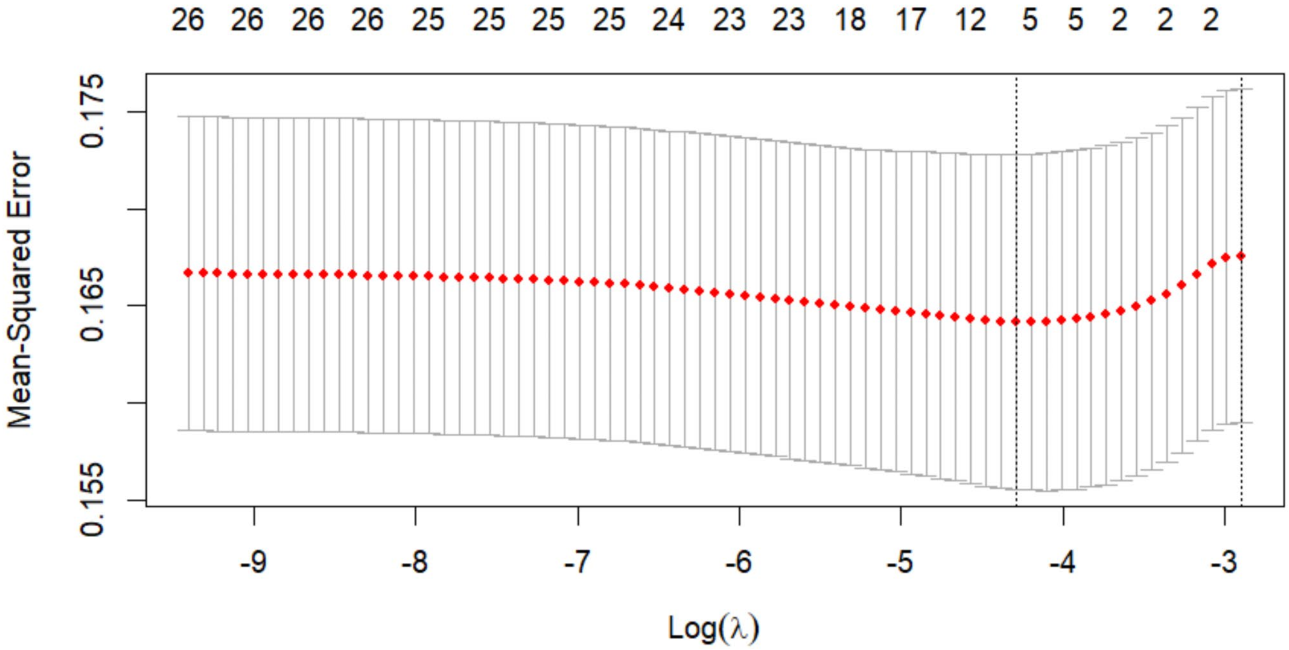

Fig. 4 The LASSO regression with 10-fold cross-validation for variate selection

values of outcome (NO ABL). It means that low-income participants are associated with ABL. Most red and blue points are concentrated on the zero line, suggesting a weak association between gender and ABL. A large proportion of points are concentrated to the left of the zero line in the race feature. The lower values of race (Mexican American and Non-Hispanic White) are associated with higher values of outcome (NO ABL). The contribution of Cd to the ML model was negative, and most of the red and blue points were on the left of the zero line. The points of SHAP in Pd are slightly concentrated to the left of the zero line and Co to the right of the zero line. The effects of Co and Pb on ABL are weaker than that of Cd. The results showed that the top three potentially critical heavy metals associated with ABL risk are Cd, Co, and Pb.

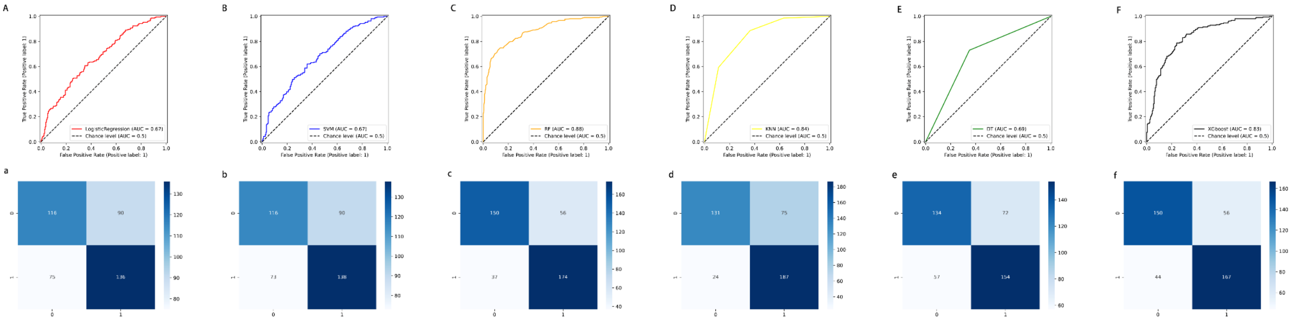

Fig. 5 ROC curve and confusion matrix for six ML models. **A**: ROC curve for LR; **B**: ROC curve for SVM; **C**: ROC curve for RF; **D**: ROC curve for KNN; **E**: ROC curve for DT; **F**: ROC curve for XGboost; **a**: confusion matrix for LR; **b**: confusion matrix for SVM; **c**: confusion matrix for RF; **d**: confusion matrix for KNN; **e**: confusion matrix for DT; **f**: confusion matrix for XGboost

Table 2 ML performance index

	LR	SVM	RF	KNN	DT	XGboost
Accuracy	0.60	0.61	0.78	0.76	0.69	0.76
AUC	0.67	0.67	0.88	0.84	0.69	0.83
Precision	0.60	0.61	0.76	0.71	0.68	0.75
Recall	0.65	0.65	0.83	0.88	0.73	0.79
F1 score	0.62	0.63	0.79	0.79	0.70	0.77

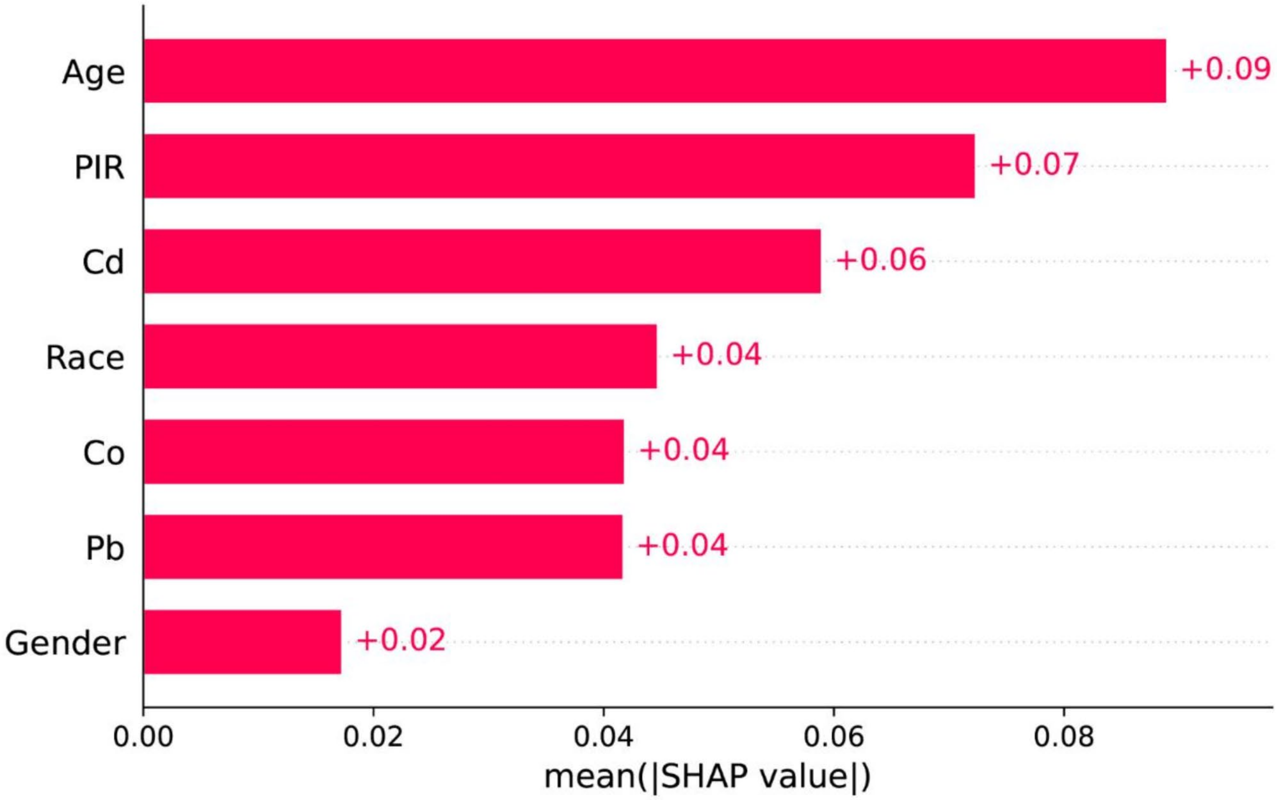

Fig. 6 The SHAP summary plot of all variables and ABL risk

Discussion

This study investigates the complex relationship between heavy metals and the ABL. Among the six ML algorithms, the RF model with seven variables showed superior performance in exploring the potential risk of ABL. The SHAP values demonstrated that Cd made the greatest contribution to the ML model among heavy metals. Age was the most significant feature among all

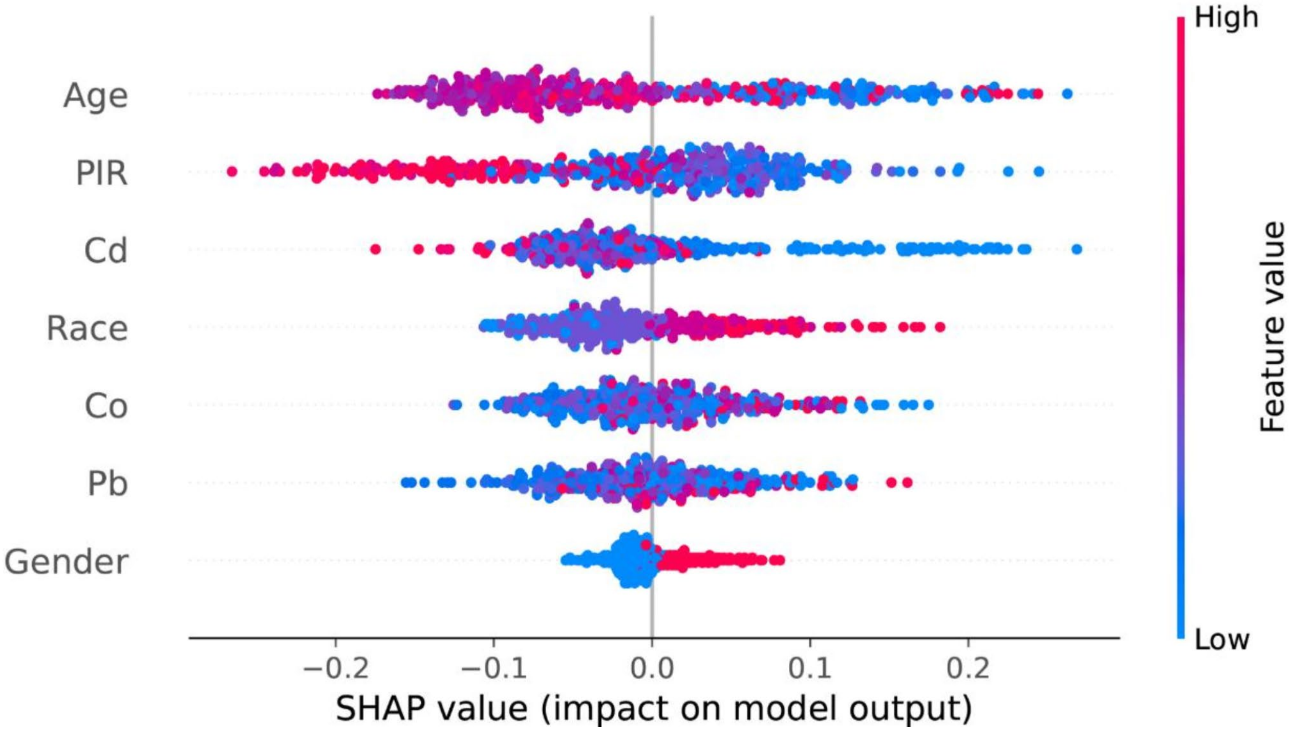

Fig. 7 The SHAP bee swarm plot of all variables and ABL risk

features, whereas sex had a minimal effect on the ML model. This model provides a theoretical basis for precision medicine research.

A systematic review reported that some researchers developed different models to evaluate periodontal bone loss based on panoramic and intraoral radiographs, and all studies demonstrated satisfactory sensitivity and specificity [36]. However, low evidence levels and inconsistent model algorithms have been presented in these studies. It was speculated that environmental factors were not considered in the prediction models. In this study, heavy metals were included as features in the construction models. Although deep learning algorithms, such as ResNet, FCN, GoogLeNet, were not used in this study, RF in ML algorithms showed superior performance of the prediction model (AUC = 0.88). RF is a multimodal algorithm based on DT. Classification and regression can be performed using the RF. The RF can assess a model with numerous features and prevent overfitting [37].

This study suggests that Cd is the most relevant to ABL among the heavy metals. A large population-based study showed the similar results that Cd concentration was positively associated with periodontitis [38]. However, this relationship might be mediated by sex hormones in above mentioned study. Exposure to heavy metals was reported to inhibit the secretion of sex hormones and affect the reproductive system [39]. Sex hormone is a risk factor for alveolar bone loss. My study revealed that gender features had little effect on the prediction model. I speculate that this difference lies in the statistical model used. The Bayesian statistical model was applied in their study, whereas the Naive Bayes algorithm could not be applied to the ML model because of the non-Gaussian distribution of the feature data. The effect of Cd on the ABL can be attributed to both direct and indirect mechanisms. Cd has a direct toxic effect on osteocytes and indirectly affects bone metabolism through regulation of blood calcium regulatory hormones [40]. A study in vitro suggested that low dose of Cd promoted the release of IL-1 and TNF-α, and IL-6 was overexpressed at a high dose of Cd. All of the above-mentioned inflammatory factors are related to periodontitis and ABL [41–42]. Another potential mechanism is that Cd causes renal damage and affects blood calcium levels and bone metabolism. An animal study revealed that kidney injury-related biomarkers were detected at high levels in animals exposed to Cd. This is evidence for the effect of Cd on bone absorption [43].

In the present study, I found that age was an important indicator of ABL. According to the literature, aging is responsible for ABL. FoxO1 transcription factor plays an indispensable role in development, senescence, cell viability, and oxidative stress in organs and cells. ABL was alleviated in FoxO1 KO mice [44]. One study suggested a strong multiple linear regression between age and ABL, and gender was not an important variable in the model [45]. This finding is consistent with the conclusions of the present study. However, race was regarded as a pivotal feature of this study. I speculate that this difference is due to the target population. All American races were included in the model in my study, while only African Americans were recruited for the study. Meanwhile, the association between PIR and ABL was non-negligible. According to an epidemiological investigation, low-income individuals are more susceptible to ABL than high-income individuals. High income is a protective factor against ABL [46]. A similar conclusion was reached. High-income individuals possess strong economic strength and complex social contacts. They can afford high medical costs and access advanced medical services if they suffer from ABL or other diseases. Whether there is an interaction between the PIR and gender or race deserves further research.

Co and Pb contributed equally to the ML model. In previous studies, Co has proven to be effective for bone formation. Cobalt-substituted hydroxyapatite can stimulate bone cell growth, alleviate inflammatory responses, and resist bacteria. Thus, it may be an optimal biomaterial for alveolar bone regeneration [47]. An in vitro study demonstrated that Co-TCP promotes ALP activity, bone matrix mineralization, and osteogenic gene expression in bone marrow stem cells. Appropriate Co combined with biomaterials has a positive effect on bone formation [48]. Whether Co combined with a bone graft can be applied for maxillary sinus floor elevation requires further basic and clinical trials. A previous animal study reported that rats exposed to Pb under hypoxic conditions were more susceptible to ABL and an increase in oral tissue inflammation parameters was detected [49]. Pb is regarded as a heavy metal and is characterized by cytotoxicity. Osteoclasts and osteoblasts were also affected. Pb disturbs protein and nucleic acid synthesis, and prevents some proteins from neutralizing free radicals [50].

Nevertheless, the present study had some limitations. First, RF demonstrated superior performance in the current model, and no other algorithms were introduced into the prediction mode (e.g., AdaBoost and neural networks). Second, this study focused on the US population and external validation requires further investigation. Further studies should investigate this relationship in Asian populations using the KNHANES database. Third, the manual selection of features might be slightly biased, and advanced algorithms can save time and cost for researchers using automated screening feature techniques. Fourth, Heavy metal exposure is often geographically dependent, yet this variable was not included in the model. Fifth, duration and dose of heavy metal exposure were not taken into account due to the limitation of database. These factors should be considered in the future research. Finally, causal associations could not be verified in this study due to lack of temporal sequencing, and the results need to be interpreted with caution.

Conclusion

In this study, a ML model was developed to explore the association between ABL risk and environmental elements combined with demographic data. RF showed superior performance compared with the other five algorithms. Among the 12 heavy metals, Cd was the most important factor in the ML model. The relationship of Co & Pb and ABL are weaker than that of Cd. Among all the independent variables, age was considered the most important factor in this model. As for PIR, low-income participants are associated with ABL. Mexican American and Non-Hispanic White have low association with ABL compared to Non-Hispanic Black and other races. Gender feature demonstrates a weak association with ABL. In the future, more advanced algorithms should be developed to validate these results and related parameters can be tuned to improve the accuracy of the model.

Acknowledgements

The author would like to thank for open data provided by NHANES.

Author contributions

Jiayi Chen made all contributions to this study.

Funding

None.

Data availability

All relevant data in this study were mentioned in the article and could be open to the public.

Declarations

Ethical approval and consent to participate

This study was conducted according to the guideline laid down in the Declaration of Helsinki, and all procedures involving study participants were approved by the Institutional Review Board of the National Center for Health Statistics (NCHS). Ethical review and approval were waived for this study as it solely used publicly available data for research and publication. Informed consent was obtained from all subjects involved in the NHANES.

Consent for publication

Not applicable.

Competing interests

The author declares no competing interests.

Author details

[1]Department of stomatology, Suzhou Wujiang District Hospital of
Traditional Chinese Medicine, Dachun road 999, Wujiang District,
Suzhou 215221, PR China

Received: 23 September 2024 / Accepted: 28 January 2025
Published online: 04 February 2025

References

1. Saffar JL, Lasfargues JJ, Cherruau M. Alveolar bone and the alveolar process: the socket that is never stable. Periodontol. 2000;13:76–90.
2. Hnitecka S, Olchowy C, Olchowy A, Dąbrowski P, Dominiak M. Advancements in alveolar bone reconstruction: A systematic review of bone block utilization in dental practice. Dent Med Probl. 2024;61:933-941.
3. Jeffcoat MK. Bone loss in the oral cavity. J Bone Min Res. 1993;8:S467–73.

4. Heinemann F, Hasan I, Bourauel C, Biffar R, Mundt T. Bone stability around dental implants: treatment related factors. Ann Anat. 2015;199:3–8.
5. Hajishengallis G. Periodontitis: from microbial immune subversion to systemic inflammation. Nat Rev Immunol. 2015;15:30–44.
6. Mainas G, Nibali L, Ide M, Mahmeed WA, Al-Rasadi K, Al-Alawi K, Banach M, et al. Associations between Periodontitis, COVID-19, and Cardiometabolic complications: Molecular mechanisms and clinical evidence. Metabolites. 2022;13:40.
7. Pirih FQ, Monajemzadeh S, Singh N,et al. Association between metabolic syndrome and periodontitis: the role of lipids, inflammatory cytokines, altered host response, and the microbiome. Periodontol 2000. 2021;87:50–75.
8. Kabasawa M, Ejiri S, Hanada K, Ozawa H. Effect of age on physiologic and mechanically stressed rat alveolar bone: a cytologic and histochemical study. Int J Adult Orthodon Orthognath Surg. 1996;11:313–27.
9. Leite FRM, Nascimento GG, Scheutz F, López R. Effect of smoking on Periodontitis: a systematic review and Meta-regression. Am J Prev Med. 2018;54:831–41.
10. Karatela S, Coomarasamy C, Paterson J, Ward NI. Household Smoking Status and Heavy Metal Concentrations in toenails of children. Int J Environ Res Public Health. 2019;16:3871.
11. Hansen BF, Gjermo P, Bellini HT, Ihanamaki K, Saxén L. Prevalence of radiographic alveolar bone loss in young adults, a multinational study. Int Dent J. 1995;45:54–61.
12. Jannetto PJ, Cowl CT. Elementary Overview of Heavy metals. Clin Chem. 2023;69:336–49.
13. World Health Organization, Fewtrell L, Kaufman R, Prüss-Üstün A. Lead: assessing the Environmental Burden of Diseases at National and local levels. Geneva, Switzerland: World Health Organization; 2003.
14. So KY, Park BH, Oh SH. Cytoplasmic sirtuin 6 translocation mediated by p62 polyubiquitination plays a critical role in cadmium-induced kidney toxicity. Cell Biol Toxicol. 2021;37:193–207.
15. Ebrahimi M, Khalili N, Razi S, Keshavarz-Fathi M, Khalili N, Rezaei N. Effects of lead and cadmium on the immune system and cancer progression. J Environ Health Sci Eng. 2020;18:335–43.
16. Buha A, Jugdaohsingh R, Matovic V, et al. Bone mineral health is sensitively related to environmental cadmium exposure- experimental and human data. Environ Res. 2019;176:108539.
17. Ma Y, Ran D, Shi X, Zhao H, Liu Z. Cadmium toxicity: a role in bone cell function and teeth development. Sci Total Environ. 2021;769:144646.
18. Karri V, Ramos D, Martinez JB, et al. Differential protein expression of hippocampal cells associated with heavy metals (Pb, As, and MeHg) neurotoxicity: Deepening into the molecular mechanism of neurodegenerative diseases. J Proteomics. 2018;187:106-25.
19. Conti MI, Terrizzi AR, Lee CM, Mandalunis PM, Bozzini C, Piñeiro AE, Martínez Mdel P. Effects of lead exposure on growth and bone biology in growing rats exposed to simulated high altitude. Bull Environ Contam Toxicol. 2012;88:1033–7.
20. Han DH, Lee HJ, Lim S. Smoking induced heavy metals and periodontitis: findings from the Korea National Health and Nutrition examination surveys 2008–2010. J Clin Periodontol. 2013;40:850–8.
21. Kwon T, Lamster IB, Levin L. Current concepts in the management of Periodontitis. Int Dent J. 2021;71:462–76.
22. Li ZH, Li J, Mao YC, et al. Association of urinary heavy metal combined exposure with periodontitis among US adults from NHANES 2011-2014. Environ Sci Pollut Res Int. 2023;30:107887-98.
23. Cho H, She J, De Marchi D, et al. Machine Learning and Health Science Research: Tutorial. J Med Internet Res. 2024;26:e50890.
24. Li S, Wen C, Bai X, Yang D. Association between biological aging and periodontitis using NHANES 2009–2014 and mendelian randomization. Sci Rep. 2024;14:10089.
25. Li W, Song J, Chen Z. The association between dietary vitamin C intake and periodontitis: result from the NHANES (2009–2014). BMC Oral Health. 2022;22:390.
26. Sanders AE, Slade GD, Fitzsimmons TR, Bartold PM. Physical activity, inflammatory biomarkers in gingival crevicular fluid and periodontitis. J Clin Periodontol. 2009;36:388–95.
27. Dormann CF, Elith J, Bacher S, et al. Collinearity: a review of methods to deal with it and a simulation study evaluating their performance[J]. Ecography. 2013;36:027–46.
28. Kim JH. Multicollinearity and misleading statistical results. Korean J Anesthesiol. 2019;72(6):558–69.
29. Lundberg S, Lee SI. A Unified Approach to Interpreting Model Predictions. Nips 2017.
30. Li X, Zhao Y, Zhang D, 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.
31. Czesnikiewicz-Guzik M, Osmenda G, Siedlinski M, et al. Causal association between periodontitis and hypertension: evidence from mendelian randomization and a randomized controlled trial of non-surgical periodontal therapy. Eur Heart J. 2019;40:3459–70.
32. Preshaw PM, Alba AL, Herrera D, et al. Periodontitis and diabetes: a two-way relationship. Diabetologia. 2012;55:21–31.
33. Walther C, Spinler K, Borof K, et al. Evidence from the Hamburg City Health Study - Association between education and periodontitis. BMC Public Health. 2022;22:1662.
34. Mondal S, Ghosh S. Gramicidin A and its complexes with cs + and tl + ions in organic solvents. A study by steady state and time resolved emission spectroscopy. J Photochem Photobiol B. 2001;60:12–24.
35. Sevcenco AM, Bevers LE, Pinkse MW, et al. Molybdenum incorporation in tungsten aldehyde oxidoreductase enzymes from Pyrococcus furiosus. J Bacteriol. 2010;192:4143–52.
36. Patil S, Joda T, Soffe B, et al. Efficacy of artificial intelligence in the detection of periodontal bone loss and classification of periodontal diseases: A systematic review. J Am Dent Assoc. 2023;154:795-804.
37. Breiman L. Random forests. Mach Learn. 2001;45:5–32.
38. Yang H, Hu X, Luo L, et al. Association of individual and combined exposures of 10 metals with periodontitis: Results from a large population-based study. J Periodontal Res. 2024;59:669-78.
39. Zeng X, Jin T, Zhou Y, Nordberg GF. Changes of serum sex hormone levels and MT mRNA expression in rats orally exposed to cadmium. Toxicology. 2003;186:109–18.
40. Schutte R, Nawrot TS, Richart T, et al. Bone resorption and environmental exposure to cadmium in women: a population study. Environ Health Perspect. 2008;116:777–83.
41. Marth E, Barth S, Jelovcan S. Influence of cadmium on the immune system. Description of stimulating reactions. Cent Eur J Public Health. 2000;8:40–4.
42. Marth E, Jelovcan S, Kleinhappl B, Gutschi A, Barth S. The effect of heavy metals on the immune system at low concentrations. Int J Occup Med Environ Health. 2001;14:375–86.
43. Prozialeck WC, VanDreel A, Ackerman CD, et al. Evaluation of cystatin C as an early biomarker of cadmium nephrotoxicity in the rat. Biometals. 2016;29:131-46.
44. Wang Z, Zhou F, Feng X, Li H, Duan C, Wu Y, Xiong Y. FoxO1/NLRP3 Inflammasome promotes Age-related alveolar bone resorption. J Dent Res. 2023;102:919–28.
45. Streckfus CF, Parsell DE, Streckfus JE, Pennington W, Johnson RB. Relationship between oral alveolar bone loss and aging among African-American and caucasian individuals. Gerontology. 1999;45:110–4.
46. Helmi MF, Huang H, Goodson JM, Hasturk H, Tavares M, Natto ZS. Prevalence of periodontitis and alveolar bone loss in a patient population at Harvard School of Dental Medicine. BMC Oral Health. 2019;19:254.
47. Lin WC, Chuang CC, Yao C, Tang CM. Effect of cobalt precursors on cobalt-hydroxyapatite used in bone regeneration and MRI. J Dent Res. 2020;99:277–84.
48. Zheng Y, Yang Y, Deng Y. Dual therapeutic cobalt-incorporated bioceramics accelerate bone tissue regeneration. Mater Sci Eng C Mater Biol Appl. 2019;99:770–82.
49. Terrizzi AR, Fernandez-Solari J, Lee CM, et al. Alveolar bone loss associated to periodontal disease in lead intoxicated rats under environmental hypoxia. Arch Oral Biol. 2013;58:1407-14.
50. Ciosek Ż, Kot K, Kosik-Bogacka D, Łanocha-Arendarczyk N, Rotter I. The effects of Calcium, Magnesium, Phosphorus, Fluoride, and lead on bone tissue. Biomolecules. 2021;11:506.

Publisher's note

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