# RESEARCH



A cross-sectional study comparing machine learning and logistic regression techniques for predicting osteoporosis in a group at high risk of cardiovascular disease among old adults

Yuyi Peng<sup>1</sup>, Chi Zhang<sup>2</sup> and Bo Zhou<sup>1\*</sup>

## Abstract

**Background** Osteoporosis has become a significant public health concern that necessitates the application of appropriate techniques to calculate disease risk. Traditional methods, such as logistic regression, have been widely used to identify risk factors and predict disease probability. However, with the advent of advanced statistics techniques, machine learning models offer promising alternatives for improving prediction accuracy. What's more, studies that use risk factors and prediction models for osteoporosis in high-risk groups for cardiovascular diseases are scarce. We aimed to explore the risk factors and disease probability of osteoporosis by comparing logistic regression with four machine learning models. By doing so, we seek to provide insights into the most effective methods for osteoporosis risk assessment and contribute to the development of tailored prevention strategies at high risk of cardiovascular disease among old adults.

**Methods** We carried out a cross-sectional investigation of a high-risk group in cardiovascular patients. A logistic regression model and four common machine learning methods,DT,RF,SVM,and XGBoost were implemented to create a prediction model using information from 211 participants who met the inclusion requirements. Metrics for calibration and discrimination were used to compare the models.

**Results** In total,211 patients were enrolled. The AUCs were 0.751 for the logistic regression model,0.72 for the SVM model,0.70 for the random forest model,0.697 for the model XGBoost, and 0.69 for the decision tree model. The logistic regression model outperforms other models for machine learning. According to the logistic regression model, there were nine predictors, including age, sex, glucose, TG (triglyceride), fracture history, stroke history, and CNV (copy number variation) nssv659422, and low-sodium salt. A well-calibrated result of 0.199 on the Brier scale. The findings of the internal validation demonstrated the high degree of repeatability of the prediction model employed in this study.

**Conclusions** In this study, we discovered that when predicting osteoporosis, a number of machine learning techniques fell short of logistic regression. In a specific population, we have innovatively developed a risk prediction model for osteoporosis events that integrates genetic and environmental factors, is an effective tool for assessing osteoporosis risk and can serve as the basis for specialized intervention approaches.

Keywords Copy number variants, Osteoporosis, Prediction model, Machine learning, Low sodium salt intake

\*Correspondence: Bo Zhou zhoubo@cmu.edu.cn Full list of author information is available at the end of the article



© 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/.

## Introduction

Osteoporosis is a systemic metabolic bone disease characterized by decreased bone mineral density(BMD), altered bone strength and increased risk of fracture. Older males and postmenopausal women are prone to this disease [1]. At present, osteoporosis is defined as one of the 10 most important diseases affecting all mankind, and the increase in global life expectancy has led to a significant increase in the proportion of people at risk of osteoporosis [2]. According to the most recent epidemiological report, 200 million people worldwide suffer from the negative impact of osteoporosis [3]. Over the previous 12 years, osteoporosis has become more common in China. In addition, the incidence of osteoporosis increases significantly with age [4]. The prevalence in people over 50 years old was 19.2%, and that in people aged 65 years was 32% [5]. Therefore, it is important to have a suitable plan in place to predict the onset of osteoporosis and provide timely intervention to prevent it.

Unlike diseases that act on a single gene, osteoporosis is the result of the interaction of multiple external and internal factors. According to Yang's research [6], there is a tight correlation between the incidence of CVD in older persons over 60 years of age and BMD. Specifically, there is a negative nonlinear relationship between the risk of cardiovascular disease and femur bone mineral density. The study population of our research article was a high-risk population for cardiovascular diseases. Zelin [7] showed that consuming calcium tablets, drinking alcohol, physical exercise, age,height,weight,body mass index,waist circumfere nce,ethnicity,menopausal status,maternal history of osteoporosis, serum estradiol history, serum DHEA levels, and serum thyrotropin levels are influential factors for osteoporosis. Our study investigated whether lowsodium diets may influence the development of osteoporosis and whether salt intake (low-sodium salt or regular salt) could be used as a predictor in an osteoporosis risk assessment model [8, 9]. Moreover, we also examined how relevant CNVs may contribute to osteoporosis. Genetic factors may be crucial for the prevention and treatment of osteoporosis [10, 11].

Many countries and regions have carried out largescale research to establish corresponding risk assessment models for osteoporosis. Several prominent risk assessment studies, including Osteoporosis Self-Screening Tool for Asians(OSTA) [12], Fracture Risk Assessment Tool (FRAX) [13] and Quantitative Ultrasound(QUS) [14] have provided extremely helpful advice for the prevention and treatment of osteoporosis [15]. To improve their predictive power, osteoporosis prediction models have been developed in recent years that focus on certain groups or incorporate new factors. For example, Lee [16] et al. created an osteoporosis prediction model for patients with rheumatoid arthritis. Wu [17] used machine learning methods, which included genetic factors, sex, number of children and breastfeeding children, age, place of residence, education level, measurement season,height,smoking status,hormone replacement therapy, serum ALB, hip circumference, vitamin B6 intake and weight, to construct a prediction model for people over 40 years old at high risk of osteoporosis. Since Bhotla [18] et al. reported that osteoporosis may be a risk factor for cardiovascular diseases, we built a relevant diagnostic model for old population undergoing a low-sodium salt intervention, hoping to conduct a better and earlier intervention for the high-risk cardiovascular population.

Predicting osteoporosis is challenging due to the complexity of the disease, which is influenced by numerous factors including genetics, lifestyle, and environmental factors. Conventional logistic regression(LR) methods have been used to construct osteoporosis prediction models [19, 20], but it may struggle to capture these complex relationships. With the advent of machine learning(ML),a data analysis method that predict outcomes by "learning" from data is becoming more and more popular in predicting osteoporosis [21, 22]. The use of SHAP techniques makes each variable interpretable and quantifiable, machine learning is no longer solely a 'black box' process, but rather one with increasing transparency and interpretability. For this study, we selected four well-established machine learning algorithms: Support Vector Machines (SVM), Random Forest(RF), Decision Tree(DT), and Extreme Gradient Boosting(XGBoost)because of their strengths in handling high-dimensional data, non-linear relationships, and their ability to identify key risk factors. These algorithms can uncover hidden patterns and potential risk factors, leading to more accurate and personalized prediction of osteoporosis risk. However, whether sophisticated machine learning algorithms can outperform traditional generalized linear models in particular domains is still under debate.

This project team has conducted a 5-year low-sodium salt intervention trial (SSaSS) in the early stage [23, 24], based on which a cross-sectional study was conducted in the later stage. This study is innovative in its focus on osteoporosis risk prediction in a high-risk CVD population, a group that has been understudied in previous research;Unlike many studies that rely solely on traditional or ML methods,we directly compared logistic regression with multiple ML models. Additionally,our model incorporates both genetic and environmental predictors, offering a more comprehensive approach to osteoporosis risk assessment. On the basis of these studies,we aimed to use machine learning and conventional logistic regression methods to develop a model incorporating data on environmental and genetic factors to predict the presence of osteoporosis and compare the results with those of previous studies.

### **Materials and methods**

#### **Ethics statements**

The institutional review board of The First Hospital of China Medical University approved the study. The license numbers are 07–1.1–01- AF-SOP-07. Every individual participant who participated in the study provided informed consent. All of the data were anonymized prior to the data.

### Patients and study design

In this study, 6 villages (4 villages in the intervention group and 2 villages in the control group) were randomly selected from the cohort who participated in the SSaSS to complete the bone mineral density measurements and field investigations before and after the osteoporosis intervention. In the previou SSaSS study, villagers in the intervention group ate low-sodium salt and those in the control group ate regular salt. Low-sodium salt means the salt substitute, contains only 75% sodium chloride, while regular salt contains 100% sodium chloride. Specifically, we chose this salt substitute was to reduce sodium intake while maintaining palatability and ensuring safety for participants. At the time of baseline survey, the salt consumption of each family was measured, and the intervention salt was distributed every three months, each time in full quantity. In addition, each time the salt was given we asked the subjects if they were eating the salt we gave them and if it was enough. These villagers are at high risk for cardiovascular diseases, which means they have a history of stroke or high blood pressure. The PASS15.0 software calculates that a minimum of 184 people are needed. After investigation, a total of 211 people were included in the study from six villages. The demographic, personal behavior and medical behavior data of the study subjects were collected through questionnaires and health examinations. Physiological and blood biochemical data related to BMD and osteoporosis were collected.

The inclusion criteria were as follows: patients with one of the following conditions were eligible for enrollment: 1 patient with a history of stroke disease(preferred enrollment) and/or 2 patients aged  $\geq$  60 years with uncontrolled hypertension (systolic blood pressure  $\geq$  140 mmHg on two measurements at the site if they were taking antihypertensive medication; systolic blood pressure  $\geq$  160 mmHg on two measurements at the site if they were not taking antihypertensive medication). The exclusion criteria were as follows: ① members who were taking potassium-preserving diuretics, ② members who were taking potassium supplements, ③ members who had severe renal impairment, ④ members whose family members could not be included in the study for other reasons, ⑤ members who seldom ate at home, and ⑥ members whose life expectancy was less than 6 months, as judged by a physician. ⑦ Family members who were already living together were included in the study.

#### **Data gathering**

The dataset was established using Epidata 3.1 software. We included the following factors in the models that may be strongly related to osteoporosis to improve accuracy and allow for broader clinical use of the models.

1. We collected relevant patient demographic data, including age, sex, physiological variables [body mass index, waist-hip ratio, blood pressure, glucose, heart rate], living habits [smoking ( $\geq 1$  cigarette/d for  $\geq 1$  year], alcohol consumption ( $\geq 100$  mL/d of Baijiu with > 50% alcohol content or  $\geq$  500 mL/d of beer for  $\geq 1$  year], milk consumption, low-sodium salt (who were in the intervention group in the SSaSS study) [23, 24], The patient's medical history includes information on medication use, heart disease, gout, dia betes, stroke, fracture, hypertension, diabetes, fracture, and history of hyperlipemia. Body mass index (BMI) was defined as body weight divided by height squared. The waist-hip ratio (WHR) was defined as the ratio of waist circumference to hip circumference. The guestionare is within the supplementary files.

2. The following laboratory test results were collected. Blood samples were collected at the Fourth Hospital of China Medical University. Low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein (HDL-C), total cholesterol (TC), and triglyceride (TG) levels were measured, The blood samples were tested for CNV.

3. The selection of CNV sites was conducted through a comprehensive review of published literature and genome-wide association studies (GWAS) focusing on bone metabolism, osteoporosis, and cardiovascular diseases. We integrated data from the Database of Genomic Variants (DGV, http://dgv.tcag.ca/dgv/ app/home) and the UCSC Genome Browser (http:// genome.ucsc.edu/) to identify relevant CNVs. The CNVs were prioritized based on the following criteria: (1) previously reported associations with bone mineral density (BMD) or osteoporosis in at least two independent studies; (2) located within or in close proximity to genes functionally implicated in bone metabolism, particularly within exon regions; (3) exhibiting a minor allele frequency (MAF) in the general population sufficient to ensure statistical power; and (4) documented in the ClinVar database with evidence supporting their clinical relevance and phenotypic correlations. This rigorous selection process ensured that the identified CNVs were both biologically plausible and statistically robust for subsequent analyses.

#### **Definition of outcomes**

The BMD (g/cm<sup>2</sup>) of the study subjects was examined using dual-energy X-ray absorptiometry (DXA). For the purpose of our study, osteoporosis was defined as a femoral neck BMD equal to or less than 2.5 standard deviations (SDs) below the mean of a young adult reference group according to the guidelines of the World Health Organization (WHO) [25].

### Statistical analysis

Participants in the osteoporosis and nonosteoporosis groups were subjected to statistical analysis. The interquartile range (IQR) and median were used to define continuous variables that did not fit the distribution. While the mean±standard deviation was used to characterize continuous variables with a normal distribution. Percentages were used for the statistical description of categorical variables. After multicollinearity diagnosis, all indicators analysed in both univariable and multivariable models [26], and the forward and backward stepwise methods with AIC minimization were used to select the variables that were eventually included in the model.

The data were fitted using a logistic regression(LR) model, and a prediction model including genetic and environmental information was created. The prediction model was assessed in terms of calibration and discrimination. The ability of the prediction model to discriminate between patients was assessed using the area under the curve (AUC). The calibration of the prediction model was assessed using the Brier score and calibration plot [27]. We used the original dataset with bootstrap resampling to assess the performance using the C-statistic and calibration plot. Apparent performance was measured over the original sample used for model development before adjustment for bias due to model overfitting. Bootstrap performance was measured by 1000 bootstrap resamples with replacement over the original sample. Specifically, we randomly selected a sample of 211 subjects with replacement among the original cohort, measured the performance based on this random sample, and repeated the resampling and performance evaluation 1000 times to arrive at a best estimate. Test performance was defined as the performance of the models from the bootstrap samples when applied to the original sample. Expected optimism used to quantify bias due to overfitting was calculated as the difference between the bootstrap performance and test performance. The optimism-corrected performance (bias-corrected performance) was calculated as the apparent performance minus the expected optimism [28]. For every threshold probability, the net benefits were displayed using decision curve analysis (DCA). The nomogram tool was used to visualize the model.

Four distinct model types were used to create machine learning models: decision tree (DT), random forest (RF), support vector machine (SVM), and extreme gradient boosting (XGBoost). In two datasets, 20% and 80% split rates were randomly selected from among the patients. Twenty percent of the patients were then used for testing. The parameters of the four machine models are shown in supplementary table S1. The model guality of classification problems can be evaluated based on several performance metrics of the confusion matrix, including true positives (TPs), true negatives (TNs), false positives (TNs), and true positives (TPs). FP, false negative rate (FN), accuracy rate, subject characteristic area under the working curve (AUC), accuracy rate, recall rate and F1 score, where accuracy rate = TP/(TP + FP), recall rate = TP/(TP + FN),  $F1 = (2 \times accuracy \times recall)/$ (accuracy+recall). The above indices are comprehensively evaluated for each machine learning model, and the model with the best performance is selected as the final differential diagnosis model. Feature importance was interpreted using Shapley Additive Explanations (SHAP). The improvement in classification performance of the new model compared to the old model was evaluated using the Net Reclassification Improvement (NRI) and the Integrated Discrimination Improvement (IDI).

#### **Experimental materials and methods**

- (1) Experimental materials
  - (DBlood genomic DNA rapid extraction kit (Sangon, B518233);(2)DNA extraction kit (B518764) for deep processing products with magnetic bead method;(3)Blood genome DNA extraction kit with magnetic bead method (B618763); (4)UNIQ-10 TRIzol total RNA extraction kit (Shengong Bio, B511321);(5)Maxima Reverse Transcriptase (Thermo Scientific, EP0743);(6)Agarose B (BBI, A600014);(7)4S Red Plus nucleic acid stain (10,000X aqueous solution) (BBI, A606695);(6)GeneRuler DNA Ladder Mix (Thermo Scientific, B300721);(9)10×PCR Buffer (without Mg<sup>2+</sup>:100 mM Tris–HCl pH 8.8 at

25°C;<sup>(iiii)</sup>500 mM KCl, 0.8%(v/v)Nonidet) (Biological, B511321);<sup>(iiii)</sup>Maxima Reverse Transcriptase (Thermo Scientific, EP0743); <sup>(iii)</sup>Agarose B (BBI, A600014); <sup>(iii)</sup>4S Red Plus nucleic acid stain (10,000X aqueous solution, BBI, A600014);

(2) Experimental method

① The reaction solution was prepared in accordance with the manufacturer's instructions.

② The solution was centrifuged as soon as the bubbles were eliminated.

③ A digital PCR system called the QX200 Droplet was utilized to identify and examine the outcomes. The following reaction temperature and duration were used.

## Results

## Statistical analysis process

The analysis of this study is shown in the supplementary figureS1. After collecting the data, we built machine learning and logistic regression models respectively to compare their performance. The strategies for building machine learning and traditional logistic regression models are different. Consider the question of sample size, when we built machine learning we grouped them in an 8:2 ratio.

#### Participants characteristic

A total of 211 participants were included in the study. The general condition, physiological indicators, psychological condition, behavior, and lifestyle of the participants are presented in Table 1. The average age of the participants, including 92 men (43.60%) and 119 women (56.40%), was  $66.04 \pm 8.60$  years. Among these 91 osteoporosis patients, 120 did not have osteoporosis.

The CNV frequency in the groups with and without osteoporosis is displayed in Table 2. A total of 219 CNVs associated with osteoporosis were found; 8 of these were found in previous studies, and 211 were found in official experiments. Of them, 134 instances of nssv659422 were duplicates, while the remaining 77 instances (either one or three copies) had variations in copy number.

#### **Risk factor selection**

All indicators were analyzed using univariate logistic regression and multifactor logistic analysis, as shown in Table 3. Because the univariate analysis in this study showed that there were fewer risk factors, if only the statistically significant factors in the univariate analysis were included as independent variables in the multivariate logistic regression analysis, important risk factors would likely be missed. Therefore, this study is to include all variables for analysis and screen meaningful variables. There were statistically significant differences in age, sex, BMI, heart rate, and alcohol consumption according to the univariate analysis. Multivariate analysis revealed that sex, low-sodium salt, Tg, stroke history, and CNV nssv659422 were significantly different (P<0.05).

# Development and validation of a nomogram with the logistic regression model

Forward and backward stepwise methods with AIC minimization were used to select the variables that were eventually included in the model. Table 4 displays the outcomes of the multivariate logistic regression analysis. The prediction model included age, sex,glucose, hyper-glycemia status,cholestral, heart rate,fracture history, stroke history,CNV nssv659422, low-sodium salt. Among them, the osteoporosis prediction model had independent risk factors for aging, sex, cholestral, and low-sodium salt. (P<0.05).We named this newly developed model Bonepredict.

The logistic model equation is:

For clinical use, we developed a nomogram model to predict osteoporosis as shown in Fig. 1. The points for the risk of osteoporosis, glucose, fracture history, CNV nssv659422, heart rate, stroke history, Cholestral,age (years),low-sodium salt,sex,and total points were arranged from top to bottom on the abscissa. At the same time, the situation of the first observation object in the data set is displayed on the graph and marked with red dots, and its corresponding points are also marked with red dots. The variables glucose=7.4, fracture history=no, CNV nssv659422=no, heart rate=62, stroke history=yes, cholestral=0.89, age=69, low sodium salt=yes, sex=male. After adding the corresponding points, the total points=509, the corresponding probability is 0.311.

As shown in the Fig. 2, the area under the auc is 0.751(0.6859–0.8166.)Youden's index was 1.399, the sensitivity was 0.82, the specificity was 0.58, and the ROC curve cutoff value was 0.363.

An appropriate model calibration was shown by the calibration plot, which showed that the anticipated probability and the actual probability were similar (Fig. 3). The model's Brier score was 0.199, which is smaller than 0.25. The results showed that the model had good calibration degree and prediction consistency.

The old adults might obtain a good net benefit from the prediction model, according to the decision curve analysis(DCA) [29]. For example, it should be noted that

## Table 1 Characteristics of the participants

Characteristics	Total(N=211)	With osteoporosis(N=91)	Without osteoporosis(N=120)	
Age[years,M <sup>a</sup> (IQR <sup>b</sup> )]*	68(11.00)	70(3.00)	66 (12.00)	
Gender*				
Male (n, %)	92(43.60%)	26 (28.60%)	66 (55.00%)	
Female (n, %)	119(56.40%)	65 (71.40%)	54 (45.00%)	
BMI <sup>c</sup> [kg/m <sup>2</sup> , M <sup>a</sup> (IQR <sup>b</sup> )] *	32.00(6.03)	33.56(6.55)	31.22 (5.46)	
WHR <sup>d</sup> [M <sup>a</sup> (IQR <sup>b</sup> )]	0.88 (0.07)	0.87(0.09)	0.88(0.06)	
Blood pressure				
$SBP^{e}[mm HG, x \pm s]$	167.867±1.72	168.74± 2.67	167.20± 2.25	
DBP <sup>f</sup> [mm HG, x ± s]	97.043±0.87	97.21±1.25	96.91±1.20	
Heart rate [times/minute,M <sup>a</sup> (IQR <sup>b</sup> )]*	78.50(17.50)	82.00(17.50)	76.25 (18.75)	
Glucose [(mmol/L, M <sup>a</sup> (IQR <sup>b</sup> )]	6.50(2.70)	6.60 (3.10)	6.45 (2.10)	
Medical history				
Fracture (n, %)	27(12.80%)	15(16.50%)	12 (10.00%)	
Hypertension (n, %)	131(62.08%)	57(62.60%)	74 (61.70%)	
Gout (n, %)	38(18.01%)	16(17.60%)	22 (18.30%)	
Diabetes (n, %)	47(22.27%)	24(26.40%)	23 (19.20%)	
Stroke (n, %)	50(23.70%)	24.(26.40%)	26 (21.70%)	
Hyperlipemia (n, %)	8(37.91%)	5(5.50%)	3 (2.50%)	
Heart diseases (n, %)	66(31.28%)	32(35.20%)	34 (28.30%)	
Living habit				
Smoking (n, %)	68(32.22%)	25(27.50%)	43 (35.80%)	
Drinking alcohol (n, %) *	58(27.49%)	18(19.80%)	40 (33.40%)	
Drinking milk (n, %)	60(28.43%)	24(26.40%)	36 (29.90%)	
Low-sodium salt (n, %)	135(63.98%)	54(59.30%)	81(67.50%)	
Medication use (n, %)	54(25.60%)	26(28.60%)	28(23.30%)	
Blood lipid parameter				
TG <sup>g</sup> [mmol/L, <b>M <sup>a</sup> (IQR <sup>b</sup> )</b> ]	1.94(1.76)	1.82(1.99)	1.97 (1.54)	
TC <sup>h</sup> [mmol/L, <b>M <sup>a</sup> (IQR <sup>b</sup> )</b> ]	4.94(1.39)	4.91(1.13)	5.00(1.51)	
HDL-C <sup>i</sup> [mmol/L, <b>M <sup>a</sup> (IQR <sup>b</sup> )</b> ]	1.33(0.56)	1.25(0.58)	1.27 (0.56)	
LDL-C <sup>j</sup> [mmol/L <b>M <sup>a</sup> (IQR <sup>b</sup> )</b> ]	3.12(0.64)	3.16(1.21)	3.09 (0.69)	
CNV <sup>k</sup> nssv659422(n, %)	77(36.49%)	27(29.67%)	50(41.67%)	

<sup>a</sup> Median

<sup>b</sup> Interquartile range

<sup>c</sup> Body mass index

<sup>d</sup> waist-to-hip ratio

<sup>e</sup> Systolic blood pressure

<sup>f</sup> Diastolic blood pressure

<sup>g</sup> Triglyceride

<sup>h</sup> Total cholesterol

<sup>i</sup> High-density lipoprotein cholesterol

<sup>j</sup> Low-density lipoprotein cholesterol

<sup>k</sup> Copy number variation

\*p < 0.05

the logistic regression model shows clinical utility over a range of risk thresholds, when the prediction probability threshold of the nomogram model is 0–0.75,particularly between 0.05 and 0.60, as it provides a greater net benefit compared to treating all or none of the patients, suggesting that the nomogram model has good clinical applicability (Fig. 4).

Clinical impact curve (CIC) analysis was performed, as shown in Fig. 5, to evaluate the clinical applicability of the risk prediction nomogram. For example, at the X-axis Variable

 
 Table 2
 CNV frequency in the osteoporosis and nonosteoporosis groups

Number of copies	CNV nssv659422				
	1 copy (%)	2 copies (%)	3 copies (%)		
Osteoporosis group (%)	3(3.30%)	64(70.33%)	24(26.37%)		
Non-osteoporosis group (%)	9(7.50%)	70(58.33%)	41(34.16%)		

risk threshold of 0.4, the number of people at risk with the clinical prediction model we constructed is about 520, and the actual number of people at risk is about 300, with a loss-benefit ratio of 75%. When the threshold probability is greater than 65% of the prediction score probability, the prediction model determines that the high-risk group of osteoporosis is highly matched with the actual group of osteoporosis, which confirms the high clinical utility of the prediction model. Combind DCA and CIC, when the threshold is between 0.65–0.75, the prediction model is not only has a good benifit but also has a effectiveness.

To validate the model internally, bootstrap resampling was used, with an AUC of 0.75(Mean of 1000 bootstrap samples). The internal validation findings show that the prediction model applied in this study has good robustness and consistency. The results of the bootstrap resampling are displayed in Fig. 6 and Table 5. Using 500 bootstrap samples over the full dataset, the expected optimism of the model was 0.014 for the c-statistic and 0.185 for the calibration slope. The apparent c-statistic for Osteoporosis was 0.751. The optimism-corrected c-statistic was 0.737, the calibration slope was 0.815, and the calibration-in-the-large for all performance measures was < 0.001, all with standard error < 0.001.

**Multivariate analysis** 

Table 3 Univariate analysis and multivariate analysis of risk factors for participants

Univariate analysis

	OR	95%CI	P-value	OR	P-value	95%Cl
Age	1.04	[1.01,1.08]	0.019	1.05	0.06	[1,1.1]
Gender	0.33	[0.18,0.58]	< 0.001	0.26	0.011	[0.09,0.74]
BMI	1.09	[1.02,1.17]	0.008	1.03	0.657	[0.91,1.15]
WHR	0.22	[0,19.53]	0.507	0.14	0.566	[0,109.93]
Blood pressure						
SBP	1	[0.99,1.01]	0.658	1.01	0.245	[0.99,1.03]
DBP	1	[0.98,1.02]	0.868	0.99	0.788	[0.96,1.03]
Heart rate	1.03	[1,1.05]	0.019	1.03	0.056	[1,1.06]
Glucose	1.07	[0.98,1.16]	0.144	1.07	0.247	[0.95,1.21]
Living habit						
Smoking	0.68	[0.38,1.23]	0.199	1.30	0.499	[0.6,2.81]
Drinking alcohol	0.49	[0.26,0.94]	0.03	0.68	0.37	[0.29,1.58]
Drinking milk	0.84	[0.45,1.54]	0.563	0.73	0.399	[0.36,1.51]
Low-sodium Salt	0.7	[0.4,1.24]	0.222	0.34	0.006	[0.15,0.73]
Medication use	1.31	[0.71,2.45]	0.388	1.08	0.842	[0.49,2.39]
Blood lipid parameter						
TG	0.85	[0.64,1.12]	0.241			
TC	0.9	[0.79,1.04]	0.145	0.78	0.022	[0.63,0.96]
HDL-C	0.94	[0.53,1.66]	0.83	0.68	0.397	[0.28,1.66]
LDL-C	1.07	[0.8,1.43]	0.645	1.11	0.562	[0.78,1.58]
CNV nssv659422	0.59	[0.33,1.05]	0.074	0.49	0.046	[0.25,0.99]
Medical history						
Fracture	1.78	[0.79,4.01]	0.166	2.36	0.089	[0,88,6.31]
Hypertension	1.04	[0.59,1.83]	0.886	0.73	0.417	[0.35,1.55]
Gout	2.27	[0.53,9.75]	0.271	3.31	0.185	[0.56,19.47]
Diabetes	1.34	[0.69,2.6]	0.38	1.01	0.982	[0.41,2.51]
Stroke	1.3	[0.68,2.45]	0.426	2.38	0.041	[1.04,5.47]
Hyperlipemia	0.95	[0.47,1.93]	0.888	1.08	0.872	[0.41,2.84]
Cardiovascular	1.04	[0.59,1.83]	0.886	0.73	0.417	[0.35,1.55]

Among them, there was multicollinearity between triglyceride and other variables, which was not included in the multi-factor analysis

Table 4 The predictors included in the logistic regression model

Variable	β	Wald	Odds ratio(95%CI)	P-value
Age*	0.048	5.416	1.049(1.008–1.092)	0.020
Sex*	1.328	15.366	3.773(1.942–7.328)	< 0.001
Glucose	0.072	1.951	1.074(0.972–1.188)	0.162
Cholestral*	-0.173	3.848	0.841(0.707-1.000)	0.050
Heart rate	0.025	3.490	1.025(0.999–1.052)	0.062
CNV nssv659422	-0.614	3.296	0.541(0.279–1.050)	0.069
Stroke history*	0.767	3.855	2.153(1.001-4.631)	0.050
Fracture history	0.843	3.188	2.322(0.921–5.856)	0.074
Low-sodium Salt *	-1.084	8.830	0.338(0.166–0.692)	0.003

\*p < 0.05

#### Performances of different models

Tables 6 and 7 which presents a summary of the data detailing the models used to predict osteoporosis, displays the specific outcomes of the various models. Among them, the SVM model performed with the highest accuracy (0.754), followed by the XGBoost (0.721),DT ranked the lowest (0.651). The XGBoost model performed well with respect to precision (0.667), followed by LR (0.600), SVM (0.560), DT(0.530),RF(0.523). The Logistic regression(LR)models showed the greatest sensitivity, with a value of 0.751, and the XGBoost model showed the lowest sensitivity (0.588). Considering that precision and sensitivity are often contradictory, we calculated the F score, an evaluation indicator that weighed precision and sensitivity. The top 3 F score models were DT(0.73), LR(0.69), and RF (0.63).

Among these statistics is the area under the curve (AUC), which is a common metric used to evaluate the performance of a diagnostic model. It measures the model's ability to distinguish between individuals with and without the condition. AUC can be interpreted as the probability that a randomly selected positive case will have a higher predicted probability score than a randomly selected negative case. The AUC illustrated that logistic regression had the best predictive performance, with AUC values of 0.751. The details are shown in Table 6. Overall, the best-performing model was Logistic regression, followed by SVM and RF. The AUCs of the different machine learning models were shown in supplementary figureS3.

Supplementary figure S2 shows the SHAP plots for four machine learning models. Although the nine most important variables of the four models are not exactly the same, some similarities can be found with logistic regression models, sex, age, glucose, heart rate and fracture history are all important predictors. In the XGBoost model, there are BMI and DBP variables; In the RF model, there are BMI, DBP, WHR, drink variables. In the support vector machine model, there are SBP, heart diseases, hypertension diseases variables. In the DT model,there are BMI,HDL-C,LDL-C,SBP variables. These important predictors differ from logistic regression models. What's more, in support vector machine models, CNV nssv659422 and low sodium salt are both important predictors. The SHAP plots show that lower levels of these two predictors (blue dots) were associated with a lower probability of Osteoporosis(SHAP value < 0), these are also the two most important factors to explore in this cross-sectional survey.



Fig. 1 Nomogram of the logistic regression model



Fig. 2 ROC curve of the logistic regression model



**Fig. 3** Calibration plot of the logistic regression model. Notes:Calibration plot (excellent calibration as the observed outcomes are close to the 45° line). Apparent performance = final prediction model on the full data; bias-corrected performance = apparent performance minus the expected optimism estimated from the bootstrap samples

Combined AUC and sensitivity, we ultimately selected Bonepredict, a model developed using logistic regression, as our final predictive tool. To assess the clinical utility of Bonepredict, we compared its performance with three widely used osteoporosis risk assessment tools: the Osteoporosis Self-Assessment Tool (OSTA), the Fracture Risk Assessment Tool (FRAX), and Quantitative Ultrasound (QUS). Bonepredict demonstrated superior discriminative ability, achieving a higher AUC in the development cohort compared to OSTA (AUC=0.62), FRAX for major osteoporotic fractures (AUC=0.71), and QUS (AUC=0.68). Furthermore, Bonepredict exhibited improved calibration and significantly better net reclassification improvement (NRI) and integrated discrimination improvement (IDI) when compared to these existing tools.



Fig. 4 DCA curve of the logistic regression model



Fig. 5 Clinical impact curve of the logistic regression model

## Discussion

Low sodium salt is the main focus of studies on salt interventions aimed at lowering hypertension. Nevertheless, there is no conclusive evidence that low sodium salt will lower the risk of osteoporosis. A prospective cohort study by Martin O'Donnell [30] involving participants



Fig. 6 The ROC curve after internal validation using bootstrap resampling(times = 1000). Note: the gray shade showed the mean AUC of bootstrap samples, the red line showed the apparent AUC

Table 5 Internal validation-bootstrap results for the full data
---

	C- Statistic Mean±SE	Calibration slope Mean±SE
Apparent performance	0.751	1
Bootstrap performance	$0.756 \pm 0.0008$	1
Test performance	$0.742 \pm 0.0003$	0.815 ± 0.0066
Expected performance	0.014 <b>±0.0008</b>	0.185 ± 0.0066
Optimisim-corrected performance	0.737 <b>±0.0008</b>	0.815 ± 0.0066

The bootstrap procedure leads to estimates of the optimismcorrected performance, which is calculated as apparent performance minus optimism. Means and empirical standard errors (SE, based on standard deviations over samples) are shown. Five hundred bootstrap repetitions were used for calculation of both the mean and SE (SEB). The test performance is defined as the performance of the models from the bootstrap samples when applied to the original sample. The expected optimism was calculated as the difference between bootstrap performance and test performance. The optimism-corrected performance was defined as apparent performance minus optimism

from 17 countries, revealed a U-shaped relationship between low-sodium salt consumption and osteoporosis risk, suggesting that both excessive and insufficient intake may elevate the risk. However, Yuan [31] reported that eating more low sodium salt increased the chance of developing osteoporosis. In this study, we investigated the potential of low-sodium salt intake as a novel predictor in an osteoporosis risk assessment model. Notably, this predictor has not been previously incorporated into existing models(e.g.FRAX), highlighting the unique contribution of our research to the field.

Osteoporosis development affects individual sodium iron concentrations. SCN4A encodes the NaV1.4 channel, which acts primarily in skeletal muscle. SCN4A gene mutation mainly leads to rapid inactivation of sodium ion channels [32]. The inactivation of channels may be incomplete or slow, which will cause the continuous opening of channels and the continuous inflow of sodium ions, directly resulting in an increase in the concentration of Na<sup>+</sup> in the cell. To maintain the conservation of the intracellular charge, the concentration of Na<sup>+</sup> increases. The exchange of  $Na^+/Ca^{2+}$  and  $Na^+/H^+$  increased, resulting in an increase in the concentration of  $Ca^{2+}$  in the intracellular environment and a decrease in the concentration of  $Ca^{2+}$  in the extracellular environment, thus enhancing the function of osteoclasts and significantly improving bone absorption, indirectly promoting the occurrence of osteoporosis. Therefore, the SCN4A gene [11, 33] CNV nssv659422 was utilized for model development in this study. The mechanism diagram is shown in supplementary figure S4.

	AUC	Accuracy	Sensitivity (Recall)	specificity	precision	F1
XGBoost	69.7%	72.1%	58.8%	72.2%	66.7%	62.5%
RF	70.0%	58.1%	54.0%	79.8%	52.3%	63.0%
DT	69.0%	65.1%	75.0%	79.4%	53.0%	73.0%
SVM	72.0%	75.4%	70.5%	77.3%	56.0%	59.5%
LR	75.0%	68.0%	82.0%	58.0%	60.0%	69.0%

#### **Table 6** Comparison the parameters of models for prediction of osteoporosis

XGBoost extreme gradient boosting, RF random forest, DT Decision tree, SVM support vector machine, LR logistic regression model, AUC Area under the curve of ROC Accuracy = (TP + TN)/(TP + TN + FP + FN)

Sensitivity = TP/(TP + FN)

Specificity = TN/(TN + FP)

Precision = TP/(TP + FP)

F1 = 2\*Precision\*Recall/ (Precision + Recall)

TP = true positive

TN = true negative

FP = false positive

FN = false negative

Table 7	Performance cor	mparison of bo	onepredict with	n existing osteo	porosis ris	k assessment tool	S

Model	AUC	Brier	NRI	Р	IDI	Р
FRAX	0.71	0.212	0.027	p<0.05	0.153	p<0.05
QUS	0.68	0.223	0.036	p < 0.05	0.162	p<0.05
OSTA	0.62	0.238	0.040	p < 0.05	0.188	p<0.05
Bonepredict	0.75	0.199				

Lifelong risk factors such as genetics can aid in the early detection of high-risk individuals and prompt lifestyle changes that are essential for osteoporosis prevention and treatment. In bone mineral prediction models, SNPs, or single-nucleotide variations, have been incorporated as genetic factor predictors [34]. However, copy number variation was not included as a genetic feature in earlier prediction models. Since CNVs span a significantly greater number of nucleotides in the genome than do SNPs and because of their wide distribution throughout the human genome, it has been established that CNVs are linked to osteoporosis [33, 35-37]. Francine et al. assessed all CNVs from a meta-analysis based on GWAS for genomic regions linked to osteoporosis to determine whether they were connected to important osteoporosis variance. In this study, multivariate logistic regression analysis was used to identify and examine one CNV locus. The final prediction model incorporated the CNV nssv659422 (SCN4A gene), which was shown to significantly (P < 0.05) increase the likelihood of developing osteoporosis. Due to sample size and budgetary limitations, we were unable to identify more CNV loci; however, our results indicate that when genetic factors,

such as CNVs, are used as predictors, model prediction accuracy can be improved by fully accounting for the combined impact of genetic and nongenetic components. Unlike OSTA and FRAX, which primarily rely on clinical risk factors, Bonepredict incorporates genetic markers (CNVs), enabling a more personalized and precise risk assessment. Notably, the SHAP analysis for the SVM model revealed that both low-sodium salt intake and CNV nssv659422 are significant predictors of osteoporosis risk. This finding underscores the importance of integrating dietary and genetic factors in osteoporosis risk assessment models.

In recent years, artificial intelligence (AI) used to diagnose osteoporosis from risk factors in clinical data and proposing protocols, we created four mainstream machine learning models (DT,SVM,XGBoost and RF). SVM was chosen for its ability to maximize class separation margins and avoid overfitting, critical given our small sample size (n = 211). RF was selected for its ensemble learning approach, which enhances accuracy while managing missing data and identifying key predictors via feature importance rankings. Its stability with limited samples further justified its use. DT served as

an interpretable baseline model, providing transparent decision pathways for clinical insights despite its overfitting tendency. XGBoost was prioritized for its structured data optimization, gradient-boosting architecture (capturing predictor interactions), and regularization techniques to prevent overfitting. While other machine learning techniques, such as neural networks, could have been considered, they were less suitable for our study. Neural networks typically require larger sample sizes to achieve optimal performance and are less interpretable.

Following a comprehensive evaluation of model performance metrics, including the Area Under the Curve (AUC) and sensitivity, we selected the logistic regression model as the finnal model. Despite testing four machine learning algorithms (DT,SVM, RF, and XGBoost), logistic regression achieved the highest AUC (0.75) with only nine variables, demonstrating greater clinical practicality, particularly when integrated with nomograms for risk visualization. The superior performance of logistic regression compared to some machine learning algorithms in our study may be attributed to several factors: first, the relatively small sample size (n = 211) in our dataset may have limited the ability of more complex ML models to fully capture the underlying patterns without overfitting. Logistic regression, with its simpler structure and fewer parameters, is less prone to overfitting in such scenarios. Second, the relationships between the predictors and the outcome (osteoporosis) in our dataset may be predominantly linear or additive, which aligns well with the assumptions of logistic regression. Third, the interpretability of logistic regression allowed for more effective feature selection and model refinement, ensuring that only clinically relevant predictors were included, whereas ML models often prioritize statistical patterns over therapeutic relevance. Finally, the absence of highly interactive or non-linear effects in our dataset may have reduced the advantage of ML algorithms, which typically excel in capturing such complex relationships. Consistent with Christodoulou et al.'s findings [38], our results underscore that logistic regression often matches or exceeds ML performance in clinical prediction tasks, particularly with limited data. While ML algorithms hold theoretical promise, their advantages diminish in datasets lacking intricate interactions or large sample sizes.

A few of the benefits of this study are as follows:first, we enhanced the predictive framework by incorporating novel predictors, including copy number variations (CNVs) associated with osteoporosis and low-sodium salt intake. This integration not only improves model performance but also provides valuable insights into the combined influence of genetic and environmental factors on osteoporosis risk. Second,to obtain the estimated probability values for the incidence of osteoporosis, regression models were utilized. This approach proved more suitable for real-world scenarios and enabled treatments in persons at high risk. Third, our model was specifically tailored for populations at high risk of cardiovascular disease (CVD), addressing a critical gap in existing tools(e.g.,FRAX) that primarily target general populations. By focusing on this high-risk group, our study lays the groundwork for more personalized and effective low-sodium salt interventions. Finally,artificial intelligence used to diagnose osteoporosis from risk factors in clinical data and proposing protocols. This innovative approach paves the way for the development of AI-driven tools in osteoporosis management.

However, the study had a number of shortcomings:first, the sample size (n=211) constrained our ability to develop sex-specific risk prediction models and limited statistical power for machine learning model. It may also compromise model accuracy. Second, the model was built for a population at high risk of cardiovascular disease. It cannot, however, be generally applicable or applied to a larger population. Third, as our study is a cross-sectional analysis,we are unable to determine a causal link. Fourth, the model's performance was considered acceptable with an AUC < 80%, and the accuracy is not very high, so it still has a room to improve. Finally, we restricted our analysis to widely used machine learning algorithms (DT, RF, SVM, XGBoost), which may not fully represent the spectrum of available methods.

To address these limitations, future research should prioritize:first, multi-center longitudinal studies with repeated measurements of sodium intake (e.g.,24h urinary sodium excretion) and bone mineral density (e.g., annual DXA scans) are needed to confirm these findings and elucidate the temporal relationship between sodium intake and bone health. Second, randomized controlled trials (RCTs) could further clarify the causal effects of low-sodium interventions on bone metabolism in this population. Third, expanded datasets encompassing diverse populations, including underrepresented subgroups (e.g., severe osteoporosis cases), to enhance model generalizability. Finally, to attain better performance, additional learning should also be investigated. Advanced techniques such as synthetic minority oversampling (SMOTE) or transfer learning to mitigate biases from imbalanced or limited data.

## Conclusions

Our algorithm for predicting osteoporosis events in high-risk cardiovascular patients performed well. The model may be used to predict osteoporosis episodes by taking into account hereditary(CNV nssv659422)and environmental factors(sex,age,triglyceride levels,heart rate,stroke,history,fracture history). Its superiority may stem from its interpretability and suitability for the dataset's size,reducing overfitting risks while maintaining transparency for clinical decision-making. It can serve as the cornerstone for individualized intervention methods and aid in basic osteoporosis prevention and management. The inclusion of the genetic variant CNV nssv659422 significantly enhanced predictive accuracy, underscoring the value of combining genetic and clinical variables for risk stratification. This tool holds promise for personalized osteoporosis prevention, enabling clinicians to identify high-risk patients for targeted interventions (lifestyle modifications, early screenings).

In conclusion,a low-sodium salt logistic regression prediction model with good discrimination, calibration, and clinical utility based on sex,age,triglyceride level, heart rate, stroke history, fracture history, and CNV nssv659422 may prove to be a valuable clinical tool for predicting osteoporosis risk in patients with high cardiovascular disease. Because it provides the best performance, the logistic regression model outperforms the machine learning model. In the future, to provide risk evaluations for certain populations and validate our technique, we will need to obtain external samples in the future. We could aslo exploring additional genetic factors or integrating a wider variety of machine learning techniques. What's more, another important avenue for future research would be to conduct longitudinal studies to track changes over time.

#### Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12877-025-05840-w.

Supplementary Material 1. Supplementary Material 2. Supplementary Material 3. Supplementary Material 4. Supplementary Material 5. Supplementary Material 6.

#### Acknowledgements

Funding from the LJKMZ20221209 under Scientific Research Project from the Education Department of Liaoning Province is gratefully ackonwleged.

#### Authors' contributions

Yuyi Peng wrote the main manuscript and Chi Zhang prepared Figs 1, 2 and 3. All authors reviewed the manuscript.

#### Funding

This study was founded by Scientific Research Project from the Education Department of Liaoning Province; Study on the early warning of osteoporosis and related fracture diseases in middle-aged and elderly people based on low sodium intervention population cohort (LJKMZ20221209). ① Association between low-sodium salt intervention with trajectory of blood pressure and cardiovascular events (LJKMZ20221179). (2) Registration Research and Prediction Model Construction Based on Nonspecific Low Back Pain Population Cohort (JYTMS20230110).

#### Data availability

All data generated or analysed during this study are included in supplementary information files.

#### Code availability

The code of this study will be made available on reasonable request.

#### Declarations

**Consent for publications** 

Not applicable.

#### **Competing interests**

The authors declare no competing interests.

#### Author details

<sup>1</sup>Department of Clinical Epidemiology and Evidence-Based Medicine, The First Hospital of China Medical University, Shenyang, Liaoning, China. <sup>2</sup>Department of Orthopedics, The Fourth Affiliated Hospital, China Medical University, Shenyang, Liaoning, China.

Received: 16 July 2024 Accepted: 5 March 2025 Published online: 29 March 2025

#### References

- Qu X, Liao M, Liu W, Cai Y, Yi Q, Long J, Tan L, Deng Y, Deng H, Chen X. Loss of Wnt16 leads to skeletal deformities and downregulation of bone developmental pathway in Zebrafish. Int J Mol Sci. 2021;22(13):6673.
- Wu X, Wang F, Cai X, Wang S. Characteristics and osteogenic mechanism of glycosylated peptides-calcium chelate. Curr Res Food Sci. 2022;5:1965–75.
- Xiao PL, Cui AY, Hsu CJ, Peng R, Jiang N, Xu XH, Ma YG, Liu D, Lu HD. Global, regional prevalence, and risk factors of osteoporosis according to the World Health Organization diagnostic criteria: a systematic review and meta-analysis. Osteoporos Int. 2022;33(10):2137–53.
- Chang Q, Huang J, He L, Xi F. Simple immunosensor for ultrasensitive electrochemical determination of biomarker of the bone metabolism in human serum. Front Chem. 2022;10: 940795.
- Deng YX, He WG, Cai HJ, Jiang JH, Yang YY, Dan YR, Luo HH, Du Y, Chen L, He BC. Analysis and validation of hub genes in blood monocytes of postmenopausal osteoporosis patients. Front Endocrinol (Lausanne). 2021;12: 815245.
- Yang Y, Huang Y. Association between bone mineral density and cardiovascular disease in older adults. Front Public Health. 2023;11: 1103403.
- Wang Y, Wang L, Sun Y, Wu M, Ma Y, Yang L, Meng C, Zhong L, Hossain MA, Peng B. Prediction model for the risk of osteoporosis incorporating factors of disease history and living habits in physical examination of population in Chongqing, Southwest China: based on artificial neural network. BMC Public Health. 2021;21(1):991.
- Hong S, Choi JW, Park JS, Lee CH. The association between dietary sodium intake and osteoporosis. Sci Rep. 2022;12(1):14594.
- Vafa M, Soltani S, Zayeri F, Niroomand M, Najarzadeh A. The relationship between sodium intake and some bone minerals and osteoporosis risk assessment instrument in postmenopausal women. Med J Islam Repub Iran. 2016;30:377.
- Connolly K. The role of genomic copy number variation (CNV) in osteoporosis. Murdoch University; 2012. https://researchportal.murdoch.edu. au/esploro/outputs/991005544780007891.
- Lovšin N. Copy Number Variation and Osteoporosis. Curr Osteoporos Rep. 2023;21(2):167–72.
- Agarwal K, Cherian KE, Kapoor N, Paul TV. OSTA as a screening tool to predict osteoporosis in Indian postmenopausal women - a nationwide study. Arch Osteoporos. 2022;17(1):121.
- Wu CH, McCloskey EV, Lee JK, Itabashi A, Prince R, Yu W, Li-Yu J, Chionh SB, Zhao Y, Shin CS, et al. Consensus of official position of IOF/ISCD FRAX initiatives in Asia-Pacific region. J Clin Densitom. 2014;17(1):150–5.

- Hans D, Baim S. Quantitative ultrasound (QUS) in the management of osteoporosis and assessment of fracture risk. J Clin Densitom. 2017;20(3):322–33.
- Carey JJ, Chih-Hsing WuP, Bergin D. Risk assessment tools for osteoporosis and fractures in 2022. Best Pract Res Clin Rheumatol. 2022;36(3):101775.
- Lee C, Joo G, Shin S, Im H, Moon KW. Prediction of osteoporosis in patients with rheumatoid arthritis using machine learning. Sci Rep. 2023;13(1):21800.
- Wu X, Park S. A prediction model for osteoporosis risk using a machinelearning approach and its validation in a large cohort. J Korean Med Sci. 2023;38(21): e162.
- Kuchi Bhotla H, Meyyazhagan A, Pushparaj K, Pappuswamy M, Chaudhary A, Arumugam VA, Balasubramanian B, Ragu Varman D, Orlacchio A, Rengasamy KRR. Prevalence of cardiovascular diseases in South Asians: scrutinizing traditional risk factors and newly recognized risk factors sarcopenia and osteopenia/osteoporosis. Curr Probl Cardiol. 2024;49(1 Pt B):102071.
- Akkus Z, Camdeviren H, Celik F, Gur A, Nas K. Determination of osteoporosis risk factors using a multiple logistic regression model in postmenopausal Turkish women. Saudi Med J. 2005;26(9):1351–9.
- Zahid FM, Faisal S, Kamal S, Shahzad K, Iram S, Ahinkorah BO, Seidu AA, Rasheed A, Hagan JE. Model selection and identification of osteoporosis risk factors in women to improve their healthcare. J Healthc Eng. 2023;2023:3571769.
- Ongphiphadhanakul B, Rajatanavin R, Chailurkit L, Piaseu N, Teerarungsikul K, Sirisriro R, Komindr S, Pauvilai G. Prediction of low bone mineral density in postmenopausal women by artificial neural network model compared to logistic regression model. J Med Assoc Thai. 1997;80(8):508–15.
- Shim JG, Kim DW, Ryu KH, Cho EA, Ahn JH, Kim JI, Lee SH. Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women. Arch Osteoporos. 2020;15(1):169.
- Neal B, Wu Y, Feng X, Zhang R, Zhang Y, Shi J, Zhang J, Tian M, Huang L, Li Z, et al. Effect of salt substitution on cardiovascular events and death. N Engl J Med. 2021;385(12):1067–77.
- Wang F, Pi Y, Zhao Y, Zhang Y, Zhou B, Li Z, Sun J, Yu Y, Tian M, Yang M, et al. Effect of salt substitution on fracture-a secondary analysis of the Salt Substitute and Stroke Study (SSaSS). BMC Med. 2024;22(1):366.
- Looker AC, Orwoll ES, Johnston CC Jr, Lindsay RL, Wahner HW, Dunn WL, Calvo MS, Harris TB, Heyse SP. Prevalence of low femoral bone density in older U.S. adults from NHANES III. J Bone Miner Res. 1997;12(11):1761–8.
- Xu R. Selection for Independent variables and regression method in logistic regression: an example analysis. Chin J Evid Based Med. 2016;16(11):1360–4.
- Efthimiou O, Seo M, Chalkou K, Debray T, Egger M, Salanti G. Developing clinical prediction models: a step-by-step guide. BMJ. 2024;386: e078276.
- Riley RD, Archer L, Snell KIE, Ensor J, Dhiman P, Martin GP, Bonnett LJ, Collins GS. Evaluation of clinical prediction models (part 2): how to undertake an external validation study. BMJ. 2024;384: e074820.
- Van Calster B, Wynants L, Verbeek JFM, Verbakel JY, Christodoulou E, Vickers AJ, Roobol MJ, Steyerberg EW. Reporting and interpreting decision curve analysis: a guide for investigators. Eur Urol. 2018;74(6):796–804.
- Mente A, O'Donnell M, Yusuf S. Sodium intake and health: what should we recommend based on the current evidence? Nutrients. 2021;13(9):3232.
- Cui Y, Sun K, Xiao Y, Li X, Mo S, Yuan Y, Wang P, Yang L, Zhang R, Zhu X. High-salt diet accelerates bone loss accompanied by activation of ion channels related to kidney and bone tissue in ovariectomized rats. Ecotoxicol Environ Saf. 2022;244: 114024.
- Rannou F, Pennec JP, Morel J, Guéret G, Leschiera R, Droguet M, Gioux M, Giroux-Metges MA. Na v1.4 and Na v1.5 are modulated differently during muscle immobilization and contractile phenotype conversion. J Appl Physiol (1985). 2011;111(2):495–507.
- Costantini A, Skarp S, Kämpe A, Mäkitie RE, Pettersson M, Männikkö M, Jiao H, Taylan F, Lindstrand A, Mäkitie O. Rare copy number variants in array-based comparative genomic hybridization in early-onset skeletal fragility. Front Endocrinol (Lausanne). 2018;9:380.
- Ho-Le TP, Center JR, Eisman JA, Nguyen HT, Nguyen TV. Prediction of bone mineral density and fragility fracture by genetic profiling. J Bone Miner Res. 2017;32(2):285–93.

- Park TJ, Hwang MY, Moon S, Hwang JY, Go MJ, Kim BJ. Identification of a copy number variation on chromosome 20q13.12 associated with osteoporotic fractures in the Korean population. Genomics Inform. 2016;14(4):216–21.
- Yang TL, Chen XD, Guo Y, Lei SF, Wang JT, Zhou Q, Pan F, Chen Y, Zhang ZX, Dong SS, et al. Genome-wide copy-number-variation study identified a susceptibility gene, UGT2B17, for osteoporosis. Am J Hum Genet. 2008;83(6):663–74.
- Oei L, Hsu YH, Styrkarsdottir U, Eussen BH, de Klein A, Peters MJ, Halldorsson B, Liu CT, Alonso N, Kaptoge SK, et al. A genome-wide copy number association study of osteoporotic fractures points to the 6p25.1 locus. J Med Genet. 2014;51(2):122–31.
- Christodoulou E, Ma J, Collins GS, Steyerberg EW, Verbakel JY, Van Calster B. A systematic review shows no performance benefit of machine learning over logistic regression for clinical prediction models. J Clin Epidemiol. 2019;110:12–22.

#### **Publisher's Note**

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