Je et al. BMC Women's Health (2025) 25:146
https://doi.org/10.1186/s12905-025-03669-4

BMC Women's Health

RESEARCH Open Access

Development and evaluation of a machine learning model for osteoporosis risk prediction in Korean women

Minkyung Je[1], Seunghyeon Hwang[2], Suwon Lee[2*] and Yoona Kim[3*]

Abstract

Background The aim of this study was to develop a machine learning (ML) model for classifying osteoporosis in Korean women based on a large-scale population cohort study. This study also aimed to assess ML model performance compared with traditional osteoporosis screening tools. Furthermore, this study aimed to examine the factors influencing the risk of osteoporosis through variable importance.

Methods Data was collected from 4199 women aged 40–69 years in the baseline survey of the Ansan and Ansung cohort of the Korean Genome and Epidemiology Study. Osteoporosis was set as the dependent variable to develop ML classification models. Independent variables included 122 factors related to osteoporosis risk, such as socio-demographic characteristics, anthropometric parameters, lifestyle factors, reproductive factors, nutrient intakes, diet quality indices, medical history, medication history, family history, biochemical parameters, and genetic factors. The six classification models were developed using ML techniques, including decision tree, random forest, multilayer perceptron, support vector machine, light gradient boosting machine, and extreme gradient boosting (XGBoost). The six ML classification models were compared with two traditional osteoporosis screening tools, including the osteoporosis risk assessment instrument (ORAI) and the osteoporosis self-assessment tool (OST). The ML model performances were evaluated and compared using the confusion matrix and area under the curve (AUC) metrics. Variable importance was assessed using the XGBoost technique to investigate osteoporosis risk factors.

Results The XGBoost model showed the highest performance out of the six ML classification models, with an accuracy of 0.705, precision of 0.664, recall of 0.830, and F1 score of 0.738. Moreover, the XGBoost model showed a higher performance on AUC than ORAI and OST. Variable importance scores were identified for 69 out of the 122 variables associated with osteoporosis risk factors. Age at menopause ranked first in variable importance. Variables of arthritis, physical activities, hypertension, education level, income level; alcohol intake, potassium intake, homeostatic model assessment for insulin resistance; energy intake, vitamin C intake, gout; and dietary inflammatory index ranked in the top 20 out of the 69 variables, using the XGBoost technique.

*Correspondence:
Suwon Lee
leesuwon@gnu.ac.kr
Yoona Kim
yoona.kim@gnu.ac.kr

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

Conclusions This study found that an XGBoost model can be utilized to classify osteoporosis in Korean women. Age at menopause is a significant factor in osteoporosis risk, followed by arthritis, physical activities, hypertension, and education level.

Keywords Machine learning, Classification, Osteoporosis, Women, Korean genome and epidemiology study (KoGES)

Background

Osteoporosis is a skeletal disease characterized by systemic disorders of bone mass content and microstructure [1]. The reduction of bone mass and density in osteoporosis patients increases the risk of osteoporotic fractures that lead to high mortality [2]. It is vital to prevent osteoporosis to alleviate the social and economic burden due to osteoporotic fractures [3].

The risk of osteoporosis is rising with the increasing elderly population [4]. A recent meta-analysis of 108 studies performed across six continents showed that global prevalence of osteoporosis subjects was estimated at 19.7% [5]. The global prevalence of osteoporosis subjects aged over 50 years was approximately 2.3 times higher in women (26.0%) than in men (11.2%) [5]. According to data from the Health Insurance Review and Assessment Service, the prevalence of osteoporosis subjects in Korea was estimated at approximately 1.047 million in 2020 and 94.3% were women [6].

According to the World Health Organization (WHO), the diagnostic criterion for osteoporosis is defined as bone mineral density (BMD) at the lumbar spine or hip that is 2.5 standard deviations or more below the average BMD of healthy young adults [7]. The BMD is measured using radiology methods, including dual X-ray absorptiometry (DXA), quantitative ultrasound (QUS), and quantitative computed tomography (QCT) [8].

There are simple screening tools to help predict the risk of osteoporosis. The osteoporosis risk assessment instrument (ORAI) evaluates osteoporosis risk by considering age, weight, and previous use of hormone replacement therapy [9]. The osteoporosis self-assessment tool (OST) uses weight and age as key indicators to assess the prediction of osteoporosis risk [10]. However, osteoporosis is influenced by various risk factors beyond age, weight, and previous use of hormone replacement therapy, and simple osteoporosis screening tools have limitations including low sensitivity and specificity [2, 11].

A decrease in bone mass is caused by an imbalance in bone remodeling through various factors [12]. Osteoporosis is related to various factors such as age and gender [13], genetics [14], medical and medication histories [15], and nutrient intake [16]. In women, menopause plays an important role in a bone mass decrease. Estrogen decreases the rate of bone remodeling activation and helps maintain the stability between bone formation and resorption [17]. However, a decreased level of estrogen in postmenopausal women can increase osteoclast activity and accelerate bone loss by 3–5% per year over 5 to 10 years, thereby increasing the risk of osteoporosis [17–19].

Machine learning (ML), a branch of artificial intelligence, is a computerized process that can classify and predict data patterns through learning experience from data [20]. Several recent studies were conducted to predict risks of hypertension [21, 22], dyslipidemia [23], type 2 diabetes mellitus [24], and breast cancer [25] using ML techniques. A few studies using ML techniques conducted for osteoporosis prediction [26–28]. These studies were performed using osteoporosis risk factors such as age, anthropometric and biochemical parameters. Inui et al. [26] developed a ML model to predict osteoporosis in 2541 elderly women without DXA data using 24 variables such as body mass index (BMI) and 22 biochemical parameters. Bui et al. [27] developed an osteoporosis prediction ML model in 1951 elderly Vietnamese women with 15 variables including height, weight, 11 biochemical parameters, and geographical location. Ou Yang et al. [28] conducted a ML model to predict osteoporosis in 5982 elderly Taiwanese, with 16 variables for men and 19 variables for women including height, weight, waist circumference, history of alcohol consumption, history of smoking, 2 medical histories, 3 obstetrics and gynecology history (for women), and 8 biochemical parameters.

Studies on osteoporosis prediction using ML techniques have also been conducted in Korean women [29, 30]. Kwon et al. [29] developed a ML model to predict osteoporosis in 1431 postmenopausal Korean women utilizing Korea National Health and Nutrition Examination Survey (KNHANES) data conducted in the national, cross-sectional study. They included age, education level, 5 anthropometric parameters, 6 biochemical parameters, 4 lifestyle factors, and 3 reproductive factors. Similarly, a study by Shim et al. [30] employed an osteoporosis prediction ML model in 1792 postmenopausal Korean women using the KNHANES data. They included osteoporosis risk factors, such as age, 4 anthropometric parameters, 4 lifestyle factors, 3 reproductive factors, and 7 medical histories.

Despite these prior efforts, few studies have extensively classified and predicted osteoporosis using ML techniques in Korean women, including various factors such as socio-demographic characteristics, anthropometric parameters, diet quality indices, nutrient intakes, reproductive factors, lifestyle factors, family history, medical history, medication history, biochemical parameters, and

genetics factors based on a large-scale population cohort study.

Therefore, the aim of this study was to develop a ML model to classify osteoporosis using multiple variables related to osteoporosis in Korean women based on a large-scale population cohort study. This study also aimed to evaluate the performance of the ML models in comparison with traditional osteoporosis screening tools. Moreover, this study aimed to examine the importance of variables to clarify to what extent factors influence the risk for osteoporosis.

Methods

Study population

This study utilized baseline survey data (2001 to 2002) from the Ansan and Ansung cohort study of the Korean Genome and Epidemiology Study (KoGES) conducted by the National Institutes of Health at the Korea Disease Control and Prevention Agency [31]. The KoGES is a large-scale cohort study that collects various data on socio-demographic characteristics, anthropometric parameters, genetic factors, lifestyle factors, dietary assessment, biochemical parameters, medical history, medication history, family history, and reproductive factors and performs follow-up studies [31].

The Ansan and Ansung cohort study comprises a baseline survey conducted from 2001 to 2002 and an 8th follow-up survey. The baseline data of the Ansan and Ansung study, part of the population-based cohorts in the KoGES, included 10,030 men and women and performed biennial surveys of residents aged 40 to 69 years in Ansan (urban) and Ansung (rural) [31, 32].

We included women aged 40 to 69 years from the baseline data of the Ansan and Ansung study in KoGES, which involved 10,030 subjects. We excluded subjects with missing data on energy intake ($n = 781$), men ($n = 4451$), missing data on menopause status ($n = 28$) and missing data on single nucleotide polymorphism (SNP) ($n = 571$). Finally, 4199 subjects were included (Fig. 1).

This study was conducted following the guidelines of the Declaration of Helsinki, with all subjects providing written informed consent. This study was approved by the Institutional Review Board of Gyeongsang National University (GIRB-A22-NX-0073) and the Korean Health and Genomic study at the Korea National Institute of Health (NBK-2023-003).

Independent variables

Table 1 presents 122 independent variables in the 11 categories. We divided 11 categories associated with osteoporosis risk factors, such as socio-demographic characteristics, anthropometric parameters, lifestyle factors, nutrient intakes, diet quality indices, medical history, medication history, family history, reproductive factors, biochemical parameters, and genetic factors.

We deemed the disease occurred if the subjects responded "yes" to the following questions: "Have you been diagnosed with the disease by a doctor?" or "Have you been currently treated for the disease?" Based on studies showing the association between medical history, medication use, and osteoporosis [15, 33], we included 14 medical history variables, such as hypertension, diabetes mellitus, allergic diseases, myocardial infarction, thyroid disease, congestive heart failure, coronary artery disease, hyperlipidemia, asthma, chronic obstructive pulmonary disease, kidney disease, cerebrovascular disease, gout, and arthritis (osteoarthritis and rheumatoid arthritis). We considered subjects as having taken the medications if they answered "yes" to the following questions: "Have you been taking medication continuously?" or "Have you experienced taking medication?" We also included 13 medication history variables such as steroids, oral contraceptives, hormone replacement therapy, anticonvulsants, anticoagulants, and medications of insulin, hypertension, arthritis, thyroid, osteoporosis, stroke, asthma, and hyperlipidemia. The family history of osteoporosis was divided into parents, siblings, others, and none. Anthropometric parameters included measuring subjects' height (cm) to the nearest 0.1 cm and body weight (kg) to the nearest 0.1 kg, with subjects wearing light clothing without shoes. The BMI was calculated as weight (kg) divided by height squared (m^2). The body fat and muscle mass were assessed by bioelectrical impedance analysis (Inbody 3.0, Biospace, Seoul, Korea). The blood pressure

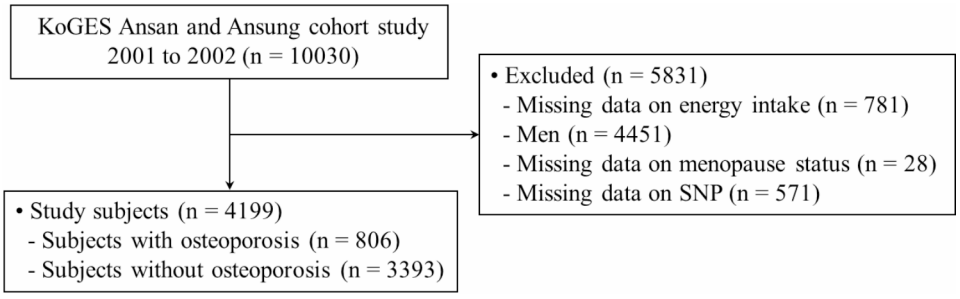

Fig. 1 A flow-diagram of the study subjects. KoGES, Korean Genome and Epidemiology Study; SNP, single nucleotide polymorphism

Table 1 Independent variables of this study

Categories	Variables	Full names
Socio-demographic characteristics	AS1_AGE	Age (year)
	AS1_EDUA	Education level
	AS1_INCOME	Income level (million Korean won/month)
Anthropometric parameters	AS1_HEIGHT	Height (cm)
	AS1_WEIGHT	Weight (kg)
	AS1_BMI	Body mass index (kg/m^2)
	AS1_WAIST	Waist circumference (cm)
	AS1_HIP	Hip circumference (cm)
	AS1_BDFTR	Body fat (%)
	AS1_BDCMSC	Muscle mass (kg)
	AS1_BPSIT1SYS	Systolic blood pressure (mmHg)
	AS1_BPSIT1DIA	Diastolic blood pressure (mmHg)
Lifestyle factors	AS1_DRINK	Alcohol intake status
	AS1_TOTALC	Alcohol intake (g/day)
	AS1_SMOKEA	Smoking status
	AS1_SLPAMTM	Average sleep time (hour)
	AS1_INSM	Insomnia
	AS1_SLPHRD	Sleep hardness
	AS1_PHYACTH	High-intensity physical activity
	AS1_PHYACTM	Moderate-intensity physical activity
	AS1_PHYACTL	Low-intensity physical activity
	AS1_PHYSIT	Sedentary physical activity
	AS1_PHYSTB	Stable physical activity
	AS1_ICOFF_1	Frequency of coffee consumption
Reproductive factors	AS1_PMYN_C	Menopausal status
	AS1_PMAG_C	Age at menopause (year)
	AS1_MNSAG	Age at menarche (year)
	AS1_PREG	Pregnancy experience status
	AS1_BRCA	Breast cancer surgery
Nutrient intakes	AS1_ENERGY	Energy intake (kcal)
	AS1_CARBO	Carbohydrate intake (g)
	AS1_PROTEIN	Protein intake (g)
	AS1_FAT	Fat intake (g)
	AS1_FIBER	Fiber intake (g)
	AS1_CALCIUM	Calcium intake (mg)
	AS1_PHOSPHO	Phosphorus intake (mg)
	AS1_MAGN	Magnesium intake (mg)
	AS1_SODIUM	Sodium intake (mg)
	AS1_POTASSIUM	Potassium intake (mg)
	AS1_FE	Iron intake (mg)
	AS1_ZN	Zinc intake (μg)
	AS1_COPPER	Copper intake (μg)
	AS1_SE	Selenium intake (μg)
	AS1_MN	Manganese intake (mg)
	AS1_CHOL	Cholesterol intake (mg)
	AS1_RETINOL	Retinol intake (μg)
	AS1_BETACARO	Beta-carotene intake (μg)
	AS1_VITD	Vitamin D intake (μg)
	AS1_VITE	Vitamin E intake (mg)
	AS1_VITK	Vitamin K intake (μg)
	AS1_VITB6	Vitamin B6 intake (mg)
	AS1_VITC	Vitamin C intake (mg)

Table 1 (continued)

Categories	Variables	Full names
Diet quality indices	AS1_INQ	Index of nutritional quality
	AS1_KHEI	Korean healthy eating index
	AS1_DII	Dietary inflammatory index
	AS1_NEAP	Net endogenous acid production
	AS1_PRAL	Potential renal acid load
	AS1_AMED	Alternate Mediterranean diet score
	AS1_DASH	Dietary Approaches to Stop Hypertension
Medical history	AS1_HT	Hypertension
	AS1_DM	Diabetes mellitus
	AS1_AL	Allergic disease
	AS1_MI	Myocardial infarction
	AS1_TH	Thyroid disease
	AS1_CH	Congestive heart failure
	AS1_CD	Coronary artery disease
	AS1_LP	Hyperlipidemia
	AS1_AS	Asthma
	AS1_CL	Chronic obstructive pulmonary disease
	AS1_KD	Kidney disease
	AS1_CV	Cerebrovascular disease
	AS1_GT	Gout
	AS1_ARRM	Arthritis (osteoarthritis and rheumatoid arthritis)
Medication history	AS1_DRST	Steroids
	AS1_DRCP	Oral contraceptives
	AS1_DRINS	Insulin medication
	AS1_DRHT	Hypertension medication
	AS1_DRAR	Arthritis medication
	AS1_DRTH	Thyroid medication
	AS1_DRFH	Hormone replacement therapy
	AS1_DROS	Osteoporosis medication
	AS1_DRSTK	Stroke medication
	AS1_DRAS	Asthma medication
	AS1_DRLP	Hyperlipidemia medication
	AS1_DRSP	Anticonvulsants
	AS1_DRSL	Anticoagulants
Family history	AS1_FMOSREL_P	Family history of osteoporosis (parents)
	AS1_FMOSREL_S	Family history of osteoporosis (siblings)
	AS1_FMOSREL_O	Family history of osteoporosis (others)
	AS1_FMOSREL_N	Family history of osteoporosis (none)

Table 1 (continued)

Categories	Variables	Full names
Biochemical parameters	AS1_ALBUMIN_TR	Albumin (g/dL)
	AS1_CREATININE_TR1	Creatinine (mg/dL)
	AS1_AST_TR	Aspartate aminotransferase (IU/L)
	AS1_ALT_TR	Alanine aminotransferase (IU/L)
	AS1_TCHL_TR	Total cholesterol (mg/dL)
	AS1_HDL_TR	High-density lipoprotein cholesterol (mg/dL)
	AS1_TOTPRT	Total protein (g/dL)
	AS1_CA	Calcium (mg/dL)
	AS1_NA	Sodium (mmol/L)
	AS1_CRP	C-reactive protein (mg/dL)
	AS1_HBA1C	Hemoglobin A1c (%)
	AS1_TG_TR	Triglycerides (mg/dL)
	AS1_WBC	White blood cell (10^3/μL)
	AS1_HCT	Hematocrit (%)
	AS1_HB	Hemoglobin (g/dL)
	AS1_BUN_TR	Blood urea nitrogen (mg/dL)
	AS1_RENIN	Renin (ng/mL/hr)
	AS1_EGFR	Estimated glomerular filtration rate (mL/min/1.73m^2)
	AS1_HOMAIR	Homeostatic model assessment for insulin resistance
Genetic factors	SNP_A-2,181,021	rs7529390
	SNP_A-1,809,518	rs628948
	SNP_A-2,130,710	rs238340
	SNP_A-2,263,153	rs6752877
	SNP_A-2,310,995	rs2722298
	SNP_A-1,922,415	rs13182402
	SNP_A-2,266,073	rs16894980
	SNP_A-1,984,271	rs3212217
	SNP_A-2,218,697	rs12100867
	SNP_A-1,850,320	rs12590815
	SNP_A-4,262,878	rs746219
	SNP_A-4,299,800	rs6064822
	SNP_A-2,242,511	rs1555364

was measured in a sitting position with the arm at heart level in a stable state. The homeostatic model assessment for insulin resistance (HOMA–IR) was calculated using fasting glucose and fasting insulin variables [34]. The estimated glomerular filtration rate (eGFR) was calculated using serum creatinine [35]. Genetic data were obtained through the Affymetrix Genome-wide Human SNP Array 5.0.

Under the categories of nutrient intakes and diet quality indices, the semi-quantitative food frequency questionnaires (SQFFQ) comprised 103 food items, assessing the frequency of each item over the past 12 months [31]. The frequency was categorized as follows: never or seldom, once a month, two or three times a month, one or two times a week, three or four times a week, five or six times a week, once a day, twice a day, or three times or more a day. Nutrient intake calculation (energy, carbohydrate, protein, fat, fiber, retinol, beta-carotene, vitamin D, vitamin E, vitamin K, vitamin B6, vitamin C, calcium, sodium, phosphorus, potassium, magnesium, iron, zinc, copper, manganese, selenium, and cholesterol) was performed with the SQFFQ data by computer aided nutritional analysis program (CAN–Pro) 5.0 software (The Korean Nutrition Society, Seoul, Korea).

Diet quality indices were calculated to estimate the impact of dietary patterns on osteoporosis including index of nutritional quality (INQ [36]), net endogenous acid production (NEAP [37]), potential renal acid load (PRAL [37]), alternate Mediterranean diet score (aMED [38, 39]), Dietary Approaches to Stop Hypertension (DASH [40, 41]), dietary inflammatory index (DII [42]), and Korean healthy eating index (KHEI [43]).

Dependent variables

Osteoporosis, the dependent variable, was evaluated based on the T-scores of the distal radius and the mid-shaft tibia BMDs using the QUS device Omnisense 7000 S/P (Sunlight Medical Ltd, Petah Tikva, Israel). We included 806 subjects with osteoporosis and 3393 subjects without osteoporosis.

Development environment and data preprocessing

The first step was data preprocessing. Missing values were imputed with the mode values for categorical variables and the mean values for continuous variables. We encoded the dataset using the categorical boosting (CatBoost) encoder, which is advantageous for handling large-scale datasets and transforming categorical and string data into continuous scalar data [44].

Osteoporosis ML classification models were implemented in Python (version 3.9.13) using libraries such as NumPy (version 1.25.2), pandas (version 1.5.3), scikit-learn (version 1.2.2), and category-encoders (version 2.6.3).

Model development and evaluation of model performance

We developed an osteoporosis ML classification model using six ML techniques, including decision tree, random forest, multi-layer perceptron (MLP), support vector machine (SVM), light gradient boosting machine (LGBM), and extreme gradient boosting (XGBoost).

Among these techniques, we primarily employed two advanced gradient boosting models, LGBM and XGBoost, as our main models while incorporating traditional ML algorithms for comparison. The selection of these models was driven by several key considerations. LGBM and XGBoost models were chosen due to their established advantages in handling structured tabular data. These models demonstrated superior performance in processing large-scale datasets with high dimensionality, which aligns well with the characteristics of our dataset. In addition, both models have exhibited robust performance on imbalanced datasets through built-in mechanisms for handling class imbalance. Their leaf-wise tree growth strategies and sophisticated regularization techniques could effectively prevent overfitting while maintaining model performance.

Meanwhile, for comparison, traditional ML algorithms such as decision tree, random forest, MLP, and SVM models were selected. These models provide a comprehensive benchmark against traditional ML methodologies, allowing us to evaluate whether the computational complexity of advanced gradient boosting techniques offers meaningful improvements over conventional approaches. These classical ML models were selected based on their distinct characteristics: decision tree for interpretability, random forest for ensemble robustness, MLP for complex non-linear relationship modeling, and SVM for effectiveness in high-dimensional spaces.

To further enhance the performance of the ML models, we conducted hyperparameter tuning was conducted using the Optuna library for model training and optimal parameter discovery. Optuna [45] is an automated software framework developed for efficient hyperparameter optimization, utilizing the Tree-structured Parzen Estimator (TPE) technique based on Bayesian Optimization. The library defines the search space through user-specified objective functions and evaluates attempted hyperparameter combinations to identify the optimal configuration.

Optuna's key features include dynamic search space definition, handling of complex constraints, and support for parallelization, enabling efficient and flexible hyperparameter optimization. These capabilities are particularly valuable in maximizing model performance when working with large-scale datasets or complex models [45].

To train and evaluate the ML models, we split the dataset into training and test datasets. The data was imbalanced with 806 subjects having osteoporosis and 3393 subjects without osteoporosis. To address this imbalance, we randomly selected 100 subjects with osteoporosis and 100 subjects without osteoporosis for the testing dataset. The remaining data was used as the training dataset. The traditional osteoporosis screening tools, ORAI and OST, were calculated for evaluation and comparison with six ML models [9, 10].

Table 2 shows the accuracy, precision, recall, and F1 score for evaluating model performance based on the confusion matrix. Accuracy measures the percentage of correctly predicted cases, indicating how well the predictions match the actual outcomes. Precision indicates the percentage of true positive instances out of the cases predicted as positive. Recall represents the proportion of correctly predicted positive instances out of all actual positive instances, reflecting the extent to which true positive outcomes were detected. The F1 score is a performance metric calculated as the harmonic mean of precision and recall.

We evaluated and compared the performance of ML classification models and traditional osteoporosis screening tools by utilizing the area under the curve (AUC) metrics. The AUC provides a comprehensive measure of a model's predictive accuracy by quantifying its ability to

Table 2 Confusion matrix and performance evaluation metrics

(a) confusion matrix

		Predicted class	
		Positive	**Negative**
Actual class	Positive	True positive	False negative
	Negative	False positive	True negative

(b) Performance evaluation metrics

Matrics	**Definition**
Accuracy	(TP + TN) / (TP + FP + FN + TN)
Precision	TP / (TP + FP)
Recall	TP / (TP + FN)
F1 score	2 × ((Recall × Precision))/(Recall + Precision)

FN, false negative; FP, false positive; TN, true negative; TP, true positive

distinguish between different classes based on the model's prediction probabilities.

The AUC serves as a single scalar value that quantifies the model's overall predictive performance. AUC values range from 0 to 1, with higher values indicating superior classification performance. Specifically, an AUC of 0.5 suggests no discriminative ability (equivalent to random guessing), while values above 0.7 are generally considered indicative of useful predictive capability. Calculation of the AUC involves analyzing the model's performance across various discrimination thresholds, considering metrics such as true positive rate (TPR, sensitivity) and false positive rate (FPR, 1-specificity).

XGBoost technique for variable importance

We used the XGBoost technique to assess the importance of various factors affecting osteoporosis. Variable importance was evaluated using 122 variables related to osteoporosis risk factors, including socio-demographic characteristics, lifestyle factors, anthropometric parameters, reproductive factors, nutrient intakes, diet quality indices, medical history, medication history, family history, biochemical parameters, and genetic factors.

Statistical analysis

The normality of distribution was assessed using the Kolmogorov-Smirnov test, Q-Q plots, and histograms. Log transformation was applied to variables that did not show normal distribution. Continuous variables were analyzed using two-sample t-test for normally distributed variables, while the Mann-Whitney U test was used for non-normally distributed variables. Categorical variables were evaluated using chi-squared analysis. Normally distributed variables are presented as means ± standard errors, while non-normally distributed variables are shown as medians and interquartile ranges. Statistical analyses were conducted using SPSS 27.0 (IBM, Chicago, IL, USA).

Results

Characteristics of study subjects

The characteristics of the study subjects are presented in Table 3. Out of a total of 4199 women, 806 subjects had osteoporosis. The osteoporosis group (median: 61 years) was older than the non-osteoporosis group (median: 49 years). The osteoporosis group (92.2%) had a higher proportion of postmenopausal women compared with the non-osteoporosis group (56.3%). The non-osteoporosis group showed significantly higher education and income levels than the osteoporosis group. The osteoporosis group had a significantly higher prevalence of hypertension, diabetes mellitus, arthritis (osteoarthritis and rheumatoid arthritis), and gout compared with the non-osteoporosis group. Moreover, the osteoporosis group had a significantly higher level of C-reactive protein (CRP) and HOMA–IR than the non-osteoporosis group. Under the categories of nutrient intakes and diet quality indices of the study subjects, the non-osteoporosis group had significantly higher intakes of energy, protein, and fat than the osteoporosis group. The osteoporosis group had significantly lower intakes of calcium, phosphorus, selenium, retinol, and beta-carotene than the non-osteoporosis group. The non-osteoporosis group consumed more vitamin B6, vitamin D, vitamin E, and vitamin K than the osteoporosis group. In diet quality indices, DII was significantly lower in the non-osteoporosis group compared with the osteoporosis group.

Model performance

Table 4 presents the performance comparison results of six ML classification models (decision tree, random forest, MLP, SVM, LGBM, and XGBoost) and two traditional osteoporosis screening tools (ORAI and OST) using ML techniques. The XGBoost model showed the highest accuracy, precision, recall, and F1 score out of the six ML classification models, with an accuracy of 0.705, precision of 0.664, recall of 0.830, and an F1 score of 0.738 (Table 4). Moreover, the XGBoost model showed a higher accuracy, precision, and F1 score than the two traditional osteoporosis screening tools. Figure 2. presents the receiver operating characteristic curves of six ML classification models and traditional osteoporosis screening tools. The six ML classification models showed a higher AUC than traditional osteoporosis screening tools. The XGBoost model had the highest performance, with an AUC of 0.84 (Fig. 2.).

Variable importance

The variable importance assessed using the XGBoost technique to evaluate factors contributing to osteoporosis risk is presented in Fig. 3. We found that 69 out of the 122 variables showed variable importance scores. Age at menopause ranked first in the variable importance. Arthritis (osteoarthritis and rheumatoid arthritis) ranked second, hypertension ranked 5th, and gout ranked 17th in the variable importance. Five physical activities ranked 3rd, 4th, 7th, 20th, and 29th in the variable importance, respectively. Renin and HOMA–IR ranked relatively high, with variable importance rankings of 11th and 14th, respectively. Education level ranked 6th, while income level and age ranked 8th and 13th in the variable importance, respectively. The variable of siblings in the family history of osteoporosis ranked 31st.

Nutrient intake variables showed relatively high importance. Potassium intake ranked 12th, while energy, vitamin C, and vitamin D intakes ranked 15th, 16th, and 18th, respectively. Moreover, diet quality indices such as DII, aMED, PRAL, and NEAP were placed at 19th, 32nd,

Table 3 Characteristics of the study subjects

Categories	Variables	Total (*n* = 4199)	Non-osteoporosis (*n* = 3393)	Osteoporosis (*n* = 806)	*P* value
Socio-demographic characteristics	Age (year)	52, 44	49, 44	61, 56	< 0.001[b]
	40–49 (n, %)	1850 (44.1)	1768 (52.1)	82 (10.2)	
	50–59 (n, %)	1118 (26.6)	872 (25.7)	246 (30.5)	
	60–69 (n, %)	1231 (29.3)	753 (22.2)	478 (59.3)	
	Education (n, %)				< 0.001
	≤ Elementary	1928 (45.9)	1338 (39.4)	590 (73.2)	
	Middle school	974 (23.2)	866 (25.5)	108 (13.4)	
	High school	1053 (25.1)	956 (28.2)	97 (12.0)	
	≥ College	244 (5.8)	233 (6.9)	11 (1.4)	
	Income level (million KRW/month)				< 0.001
	< 2	2960 (70.5)	2273 (67.0)	687 (85.2)	
	2–3.99	1000 (23.8)	905 (26.7)	95 (11.8)	
	≥ 4	239 (5.7)	215 (6.3)	24 (3.0)	
Anthropometric parameters	Height (cm)	153.8 ± 0.1	154.3 ± 0.1	151.6 ± 0.2	< 0.001[a]
	Weight (kg)	58.9 ± 0.1	58.8 ± 0.1	59.3 ± 0.3	0.435 [a]
	BMI (kg/m^2)	24.9 ± 0.0	24.7 ± 0.1	25.8 ± 0.1	< 0.001 [a]
	Waist circumference (cm)	81.7 ± 0.1	80.8 ± 0.2	85.7 ± 0.3	< 0.001 [a]
	Hip circumference (cm)	93.6 ± 0.1	93.5 ± 0.1	93.9 ± 0.2	0.201 [a]
	Body fat (%)	31.9, 28.4	31.6, 28.1	33.4, 30.1	< 0.001[b]
	Muscle mass (kg)	37.7 ± 0.1	37.9 ± 0.1	37.0 ± 0.2	< 0.001 [a]
	Blood pressure (mmHg)				
	SBP	122, 110	120, 108	130, 120	< 0.001[b]
	DBP	80, 72	80, 72	84, 78	< 0.001[b]

Table 3 (continued)

Categories	Variables	Total (*n* = 4199)	Non-osteoporosis (*n* = 3393)	Osteoporosis (*n* = 806)	*P* value
Lifestyle factors	Alcohol intake status (n, %)				< 0.001
	Current	1058 (25.2)	909 (26.8)	149 (18.5)	
	Former	129 (3.1)	105 (3.1)	24 (3.0)	
	Never	3012 (71.7)	2379 (70.1)	633 (78.5)	
	Alcohol intake (g/day)	0.0, 0.0	0.0, 0.0	0.0, 0.0	< 0.001[b]
	Smoking status (n, %)				0.155
	Current	143 (3.4)	117 (3.4)	26 (3.3)	
	Former	56 (1.3)	39 (1.1)	17 (2.1)	
	Never smoked	4000 (95.3)	3237 (95.4)	763 (94.7)	
	Average sleep time (hours)	7, 6	7, 6	7, 6	0.003[b]
	< 7 (n, %)	1880 (44.8)	1553 (45.8)	327 (40.6)	
	7–9 (n, %)	2214 (52.7)	1768 (52.1)	446 (55.3)	
	> 9 (n, %)	105 (2.5)	72 (2.1)	33 (4.1)	
	Insomnia (n, %)	896 (21.3)	664 (19.6)	232 (28.8)	< 0.001
	Sleep hardness (n, %)	709 (16.9)	520 (15.3)	189 (23.4)	< 0.001
	Physical activity (≥ 30 min/day)				
	High-intensity	1357 (32.3)	1033 (30.4)	324 (40.2)	< 0.001
	Moderate-intensity	1411 (33.6)	1176 (34.7)	235 (29.2)	0.052
	Low-intensity	3856 (91.8)	3137 (92.5)	719 (89.2)	0.020
	Sedentary	3703 (88.2)	3010 (88.7)	693 (86.0)	0.060
	Stable	2142 (51.0)	1763 (52.0)	379 (47.0)	0.014
	Coffee consumption (n, %)				< 0.001
	0 cup/week	1198 (28.5)	902 (26.6)	296 (36.7)	
	< 1 cup/week	281 (6.7)	224 (6.6)	57 (7.1)	
	1–6 cup/week	712 (17.0)	584 (17.2)	128 (15.9)	
	1 cup/day	1230 (29.3)	1014 (29.9)	216 (26.8)	
	≥ 2 cup/day	778 (18.5)	669 (19.7)	109 (13.5)	
Reproductive factors	Menopausal status (yes) (n, %)	2654 (63.2)	1911 (56.3)	743 (92.2)	< 0.001
	Age at menopause (year)	47, 45	47, 45	48, 45	0.147[b]
	Age at menarche (year)	16, 15	16, 14	16, 15	< 0.001[b]
	Pregnancy experience (yes) (n, %)	4157 (99.0)	3359 (99.0)	798 (99.0)	0.981
	Breast cancer surgery (n, %)	24 (0.6)	18 (0.5)	6 (0.7)	0.469

Table 3 (continued)

Categories	Variables	Total (*n* = 4199)	Non-osteoporosis (*n* = 3393)	Osteoporosis (*n* = 806)	*P* value
Nutrient intakes	Energy (Kcal/day)	1681.5, 1383.3	1692.3, 1395.1	1640.6, 1345.4	0.008[b]
	Carbohydrate (g/day)	301.9, 259.6	302.7, 260.1	298.9, 258.0	0.780[b]
	Protein (g/day)	55.6 ± 0.4	56.1 ± 0.4	53.3 ± 0.9	< 0.001[a]
	Fat (g/day)	28.9 ± 0.3	29.7 ± 0.3	25.6 ± 0.7	< 0.001[a]
	Fiber (g/day)	20.6 ± 0.2	20.6 ± 0.2	20.6 ± 0.4	0.244[a]
	Calcium (mg/day)	384.2, 266.0	392.5, 270.7	347.0, 243.1	< 0.001[b]
	Phosphorus (mg/day)	877.8, 661.5	887.4, 669.4	847.3, 623.3	0.002[b]
	Magnesium (mg/day)	106.0, 52.8	105.7, 53.9	108.5, 49.0	0.515[b]
	Sodium (mg/day)	2170.0 ± 20.3	2178.3 ± 22.5	2135.2 ± 47.6	0.024[a]
	Potassium (mg/day)	2637.8 ± 20.7	2637.9 ± 22.4	2637.7 ± 51.7	0.178[a]
	Iron (mg/day)	12.2 ± 0.1	12.2 ± 0.1	12.2 ± 0.2	0.365[a]
	Zinc (mg/day)	10.7, 8.4	10.8, 8.4	10.4, 8.2	0.074[b]
	Copper (μg/day)	598.3, 339.6	600.8, 341.3	587.0, 335.4	0.766[b]
	Selenium (μg/day)	35.7 ± 0.4	36.8 ± 0.4	31.0 ± 1.0	< 0.001[a]
	Manganese (mg/day)	2.1, 0.9	2.1, 0.9	2.2, 0.9	0.545[b]
	Cholesterol (mg/day)	127.2, 72.2	133.6, 77.4	98.1, 58.7	< 0.001[b]
	Retinol (μg/day)	64.6, 31.7	67.5, 34.6	48.5, 21.6	< 0.001[b]
	Beta-carotene (μg/day)	2423.6 ± 28.1	2459.1 ± 31.3	2274.0 ± 63.6	< 0.001[a]
	Vitamin D (μg/day)	1.9, 0.9	2.1, 1.0	1.4, 0.6	< 0.001[b]
	Vitamin E (mg/day)	9.9 ± 0.1	10.0 ± 0.1	9.5 ± 0.2	< 0.001[a]
	Vitamin K (μg/day)	100.1, 59.6	102.0, 62.0	90.5, 50.4	< 0.001[b]
	Vitamin B6 (mg/day)	1.8 ± 0.0	1.8 ± 0.0	1.7 ± 0.0	< 0.001[a]
	Vitamin C (mg/day)	142.9 ± 2.2	142.0 ± 2.3	146.8 ± 5.9	0.161[a]
Diet quality indices	INQ	1.0, 0.8	1.0. 0.8	1.0, 0.8	0.280[b]
	KHEI	43.6 ± 0.2	44.1 ± 0.2	41.8 ± 0.4	< 0.001[a]
	DII	2.2, 1.7	2.2, 1.7	2.4, 1.8	< 0.001[b]
	NEAP	37.7 ± 0.2	38.0 ± 0.2	36.2 ± 0.4	< 0.001[a]
	PRAL	14.8, 7.1	15.3, 7.6	13.2, 5.0	< 0.001[b]
	aMED	4, 3	4, 3	4, 3	0.928[b]
	DASH	25, 22	25, 22	25, 23	0.014[b]
Medical history	Hypertension (n, %)	759 (18.1)	537 (15.8)	222 (27.5)	< 0.001
	Diabetes mellitus (n, %)	263 (6.3)	198 (5.8)	65 (8.1)	0.019
	Allergic diseases (n, %)	275 (6.5)	236 (7.0)	39 (4.8)	0.029
	Myocardial infarction (n, %)	30 (0.7)	24 (0.7)	6 (0.7)	0.911
	Thyroid disease (n, %)	209 (5.0)	170 (5.0)	39 (4.8)	0.840
	Congestive heart failure (n, %)	9 (0.2)	7 (0.2)	2 (0.2)	0.817
	Coronary artery disease (n, %)	37 (0.9)	30 (0.9)	7 (0.9)	0.966
	Hyperlipidemia (n, %)	88 (2.1)	70 (2.1)	18 (2.2)	0.762
	Asthma (n, %)	112 (2.7)	89 (2.6)	23 (2.9)	0.715
	COPD (n, %)	18 (0.4)	17 (0.5)	1 (0.1)	0.141
	Kidney disease (n, %)	147 (3.5)	119 (3.5)	28 (3.5)	0.963
	Cerebrovascular disease (n, %)	46 (1.1)	34 (1.0)	12 (1.5)	0.233
	Gout (n, %)	391 (9.3)	279 (8.2)	112 (13.9)	< 0.001
	Arthritis (n, %)	937 (22.3)	661 (19.5)	276 (34.2)	< 0.001

Table 3 (continued)

Categories	Variables	Total (*n* = 4199)	Non-osteoporosis (*n* = 3393)	Osteoporosis (*n* = 806)	*P* value
Medication history	Steroids	14 (0.3)	13 (0.4)	1 (0.1)	0.251
	Oral contraceptives	42 (1.0)	36 (1.1)	6 (0.7)	0.417
	Insulin medication	38 (0.9)	30 (0.9)	8 (1.0)	0.770
	Hypertension medication	593 (14.1)	411 (12.1)	182 (22.6)	< 0.001
	Arthritis medication	268 (6.4)	190 (5.6)	78 (9.7)	< 0.001
	Thyroid medication	110 (2.6)	88 (2.6)	22 (2.7)	0.828
	Hormone replacement therapy	184 (4.4)	168 (5.0)	16 (2.0)	< 0.001
	Osteoporosis medication	144 (3.4)	99 (2.9)	45 (5.6)	< 0.001
	Stroke medication	11 (0.3)	7 (0.2)	4 (0.5)	0.148
	Asthma medication	48 (1.1)	40 (1.2)	8 (1.0)	0.655
	Hyperlipidemia medication	27 (0.6)	20 (0.6)	7 (0.9)	0.373
	Anticonvulsants	1 (0.0)	1 (0.0)	0 (0.0)	0.626
	Anticoagulants	1 (0.0)	0 (0.0)	1 (0.1)	0.040
Family history	Osteoporosis				
	Parents	141 (3.4)	134 (3.9)	7 (0.9)	< 0.001
	Siblings	23 (0.5)	17 (0.5)	6 (0.7)	0.400
	Other	2 (0.0)	2 (0.1)	0 (0.0)	0.491
Biochemical parameters	Albumin (g/dL)	4.1, 4.0	4.1, 4.0	4.1, 4.0	0.025[b]
	Creatinine (mg/dL)	0.7, 0.7	0.7, 0.7	0.7, 0.7	0.180[b]
	AST (IU/L)	25.0, 22.0	25.0, 22.0	26.0, 22.0	< 0.001[b]
	ALT (IU/L)	19.0, 16.0	19.0, 16.0	21.0, 17.0	< 0.001[b]
	Total-C (mg/dL)	192.2 ± 0.6	190.1 ± 0.6	201.0 ± 1.3	< 0.001 [a]
	HDL-C (mg/dL)	44.0, 39.0	45.0, 39.0	43.0, 37.0	< 0.001[b]
	Total protein (g/dL)	7.2, 7.0	7.2, 7.0	7.2, 7.0	0.367[b]
	Calcium (mg/dL)	9.6, 9.3	9.6, 9.3	9.6, 9.3	0.125[b]
	Sodium (mmol/L)	143.0, 141.0	142.0, 141.0	143.0, 142.0	< 0.001[b]
	CRP (mg/dL)	0.1, 0.1	0.1, 0.1	0.2, 0.1	< 0.001[b]
	HbA1c (%)	5.6, 5.3	5.6, 5.3	5.7, 5.5	< 0.001[b]
	TG (mg/dL)	127.0, 95.0	123.0, 93.0	143.0, 108.0	< 0.001[b]
	WBC (10^3/μL)	6.4 ± 0.0	6.4 ± 0.0	6.4 ± 0.1	0.349 [a]
	Hematocrit (%)	38.1, 36.3	38.1, 36.1	38.5, 36.9	< 0.001[b]
	Hemoglobin (g/dL)	12.6, 12.0	12.6, 12.0	12.7, 12.2	< 0.001[b]
	BUN (mg/dL)	13.8 ± 0.1	13.5 ± 0.1	14.7 ± 0.2	< 0.001 [a]
	Renin (ng/mL/hour)	2.2 ± 0.0	2.2 ± 0.0	2.2 ± 0.1	0.017 [a]
	eGFR (mL/min/1.73m^2)	202.7, 169.8	207.0, 174.6	194.3, 157.1	< 0.001[b]
	HOMA–IR	1.5, 1.1	1.5, 1.1	1.7, 1.2	< 0.001[b]

Table 3 (continued)

Categories	Variables	Total (*n* = 4199)	Non-osteoporosis (*n* = 3393)	Osteoporosis (*n* = 806)	*P* value
Genetic factors	rs7529390				0.005
	AA	58 (1.4)	41 (1.2)	17 (2.1)	
	AC	839 (20.0)	653 (19.2)	186 (23.1)	
	CC	3302 (78.6)	2699 (79.5)	603 (74.8)	
	rs628948				0.469
	AA	1216 (29.0)	973 (28.7)	243 (30.1)	
	GA	2077 (49.5)	1676 (49.4)	401 (49.8)	
	GG	906 (21.6)	744 (21.9)	162 (20.1)	
	rs238340				0.005
	GG	142 (3.4)	116 (3.4)	26 (3.2)	
	GT	1400 (33.3)	1092 (32.2)	308 (38.2)	
	TT	2657 (63.3)	2185 (64.4)	472 (58.6)	
	rs6752877				0.684
	AA	3959 (94.3)	3194 (94.1)	765 (94.9)	
	CA	235 (5.6)	195 (5.7)	40 (5.0)	
	CC	5 (0.1)	4 (0.1)	1 (0.1)	
	rs2722298				< 0.001
	CC	2706 (64.4)	2234 (65.8)	472 (58.6)	
	TC	1347 (32.1)	1042 (30.7)	305 (37.8)	
	TT	146 (3.5)	117 (3.4)	29 (3.6)	
	rs13182402				0.399
	CC	15 (0.4)	12 (0.4)	3 (0.4)	
	CT	288 (6.9)	224 (6.6)	64 (7.9)	
	TT	3896 (92.8)	3157 (93.0)	739 (91.7)	
	rs16894980				0.335
	AA	67 (1.6)	51 (1.5)	16 (2.0)	
	AG	908 (21.6)	746 (22.0)	162 (20.1)	
	GG	3224 (76.8)	2596 (76.5)	628 (77.9)	
	rs3212217				0.865
	CC	1130 (26.9)	911 (26.8)	219 (27.2)	
	GC	2138 (50.9)	1724 (50.8)	414 (51.4)	
	GG	931 (22.2)	758 (22.3)	173 (21.5)	
	rs12100867				0.235
	AA	2167 (51.6)	1768 (52.1)	399 (49.5)	
	GA	1707 (40.7)	1372 (40.4)	335 (41.6)	
	GG	325 (7.7)	253 (7.5)	72 (8.9)	
	rs12590815				0.203
	AA	118 (2.8)	102 (3.0)	16 (2.0)	
	AG	1118 (26.6)	911 (26.8)	207 (25.7)	
	GG	2963 (70.6)	2380 (70.1)	583 (72.3)	
	rs746219				0.097
	GG	1977 (47.1)	1619 (47.7)	358 (44.4)	
	TG	1790 (42.6)	1439 (42.4)	351 (43.5)	
	TT	432 (10.3)	335 (9.9)	97 (12.0)	
	rs6064822				0.916
	AA	32 (0.8)	25 (0.7)	7 (0.9)	
	AG	644 (15.3)	522 (15.4)	122 (15.1)	
	GG	3523 (83.9)	2846 (83.9)	677 (84.0)	
	rs1555364				0.970
	AA	3518 (83.8)	2841 (83.7)	677 (84.0)	
	GA	646 (15.4)	524 (15.4)	122 (15.1)	
	GG	35 (0.8)	28 (0.8)	7 (0.9)	

Table 3 (continued)

Categories	Variables	Total (*n* = 4199)	Non-osteoporosis (*n* = 3393)	Osteoporosis (*n* = 806)	*P* value

Normally distributed values are presented as means ± standard errors. Non-normally distributed values are shown as medians and interquartile ranges

Categorical variables were analyzed by chi-squared analysis

[a] Parametric values were analyzed by two-sample t-test

[b] Non-parametric values were analyzed by Mann-Whitney U test

ALT, alanine aminotransferase; aMED, alternate Mediterranean diet; AST, aspartate aminotransferase; BMI, body mass index; BUN, blood urea nitrogen; COPD, chronic obstructive pulmonary disease; CRP, C-reactive protein; DASH, Dietary Approach to Stop Hypertension; DBP, diastolic blood pressure; DII, dietary inflammatory index; eGFR, estimated glomerular filtration rate; HbA1c, hemoglobin A1c; HDL-C, high-density lipoprotein cholesterol; HOMA–IR, homeostatic model assessment for insulin resistance; INQ, index of nutritional quality; KHEI, Korean healthy eating index; KRW, Korean won; n, number; NEAP, net endogenous acid production; PRAL, potential renal acid load; SBP, systolic blood pressure; TG, triglycerides; Total-C, total cholesterol; WBC, white blood cell

Table 4 The results of classification models using machine learning techniques on validation data

Classification models	Accuracy	Precision	Recall	F1 score
Decision tree	0.560	0.597	0.370	0.457
Random forest	0.550	0.632	0.240	0.348
XGBoost	0.705	0.664	0.830	0.738
SVM	0.590	0.594	0.570	0.582
LGBM	0.635	0.642	0.610	0.626
MLP	0.620	0.607	0.680	0.642
ORAI	0.420	0.286	0.880	0.431
OST	0.670	0.357	0.400	0.377

F1 score, harmonic mean of precision and recall; LGBM, light gradient boosting machine; MLP, multilayer perceptron; ORAI, osteoporosis risk assessment instrument; OST, osteoporosis self-assessment tool; SVM, support vector machine; XGBoost, extreme gradient boosting

39th, and 62nd in the variable importance, respectively. The genetic factors of rs746219, rs12590815, rs238340, and rs628948 showed relatively low importance, ranking 41st to 44th, respectively.

Discussion

This study aimed to develop a ML model for osteoporosis classification in Korean women based on a large-scale population cohort study. We also aimed to evaluate ML model performance compared with traditional osteoporosis screening tools, including ORAI and OST. Furthermore, we aimed to investigate factors affecting osteoporosis risk in Korean women by examining variable importance.

We found that XGBoost model had the highest performance out of the 6 ML classification models, including decision tree, random forest, MLP, SVM, LGBM, and XGBoost. Moreover, we found that 6 ML classification models had higher AUC performance than 2 traditional osteoporosis screening tools, including ORAI and OST. The XGBoost model had the highest performance of AUC out of the 6 ML classification models and 2 traditional osteoporosis screening tools.

Several studies developed decision tree, logistic regression, random forest, k-nearest neighbor (KNN), SVM, neural network, artificial neural network (ANN), LGBM, and gradient boosting trees of ML models in osteoporosis risk prediction [26–28]. Inui et al. [26] employed 5 ML models including decision tree, logistic regression, random forest, LGBM, and gradient boosting trees. They found that LGBM model had the highest performance out of the five ML models [26]. Bui et al. [27] developed 4 ML models, such as logistic regression, random forest, SVM, and neural networks. They found that random forest model had the highest performance out of the 4 ML models. Oh Yang et al. [28] developed 5 ML models, including ANN, SVM, random forest, KNN, and logistic regression. They found that random forest model had the highest area under the receiver operating characteristic curve (AUROC) out of the 5 ML models [28].

Furthermore, previous studies [29, 30] have developed several ML models, including random forest, decision tree, logistic regression, gradient boosting machine, SVM, ANN, adaptive boosting (AdaBoost), and KNN, to predict osteoporosis risk in Korean women. In 3 ML models of random forest, gradient boosting machine, and AdaBoost for the prediction of osteoporosis risk with postmenopausal Korean women which developed by Kwon et al. [29], the Adaboost model showed the highest performance out of the 3 ML models. Moreover, Shim et al. [30] developed 7 ML models, including random forest, decision tree, logistic regression, gradient boosting machine, SVM, ANN, and KNN for the prediction of osteoporosis risk in postmenopausal Korean women. They found that an ANN model had the highest AUROC value out of the 7 ML models [30].

We found variable importance scores for 69 out of the 122 variables associated with osteoporosis risk factors. In our findings, age at menopause and the subject's age ranked 1st and 8th in the variable importance, respectively, aligning with previous research demonstrating the crucial role of age in osteoporosis risk. A recent prospective longitudinal study showed that women with early menopause and premature ovarian insufficiency (31.3%) had an approximately 1.43 times higher risk of osteoporosis compared with women with usual age at menopause (21.8%) [46]. In a cross-sectional study with 2224 Chinese women aged 40 to 80 years, the association between earlier menopause and the prevalence of osteoporosis was observed [47]. A reduction in estrogen levels after

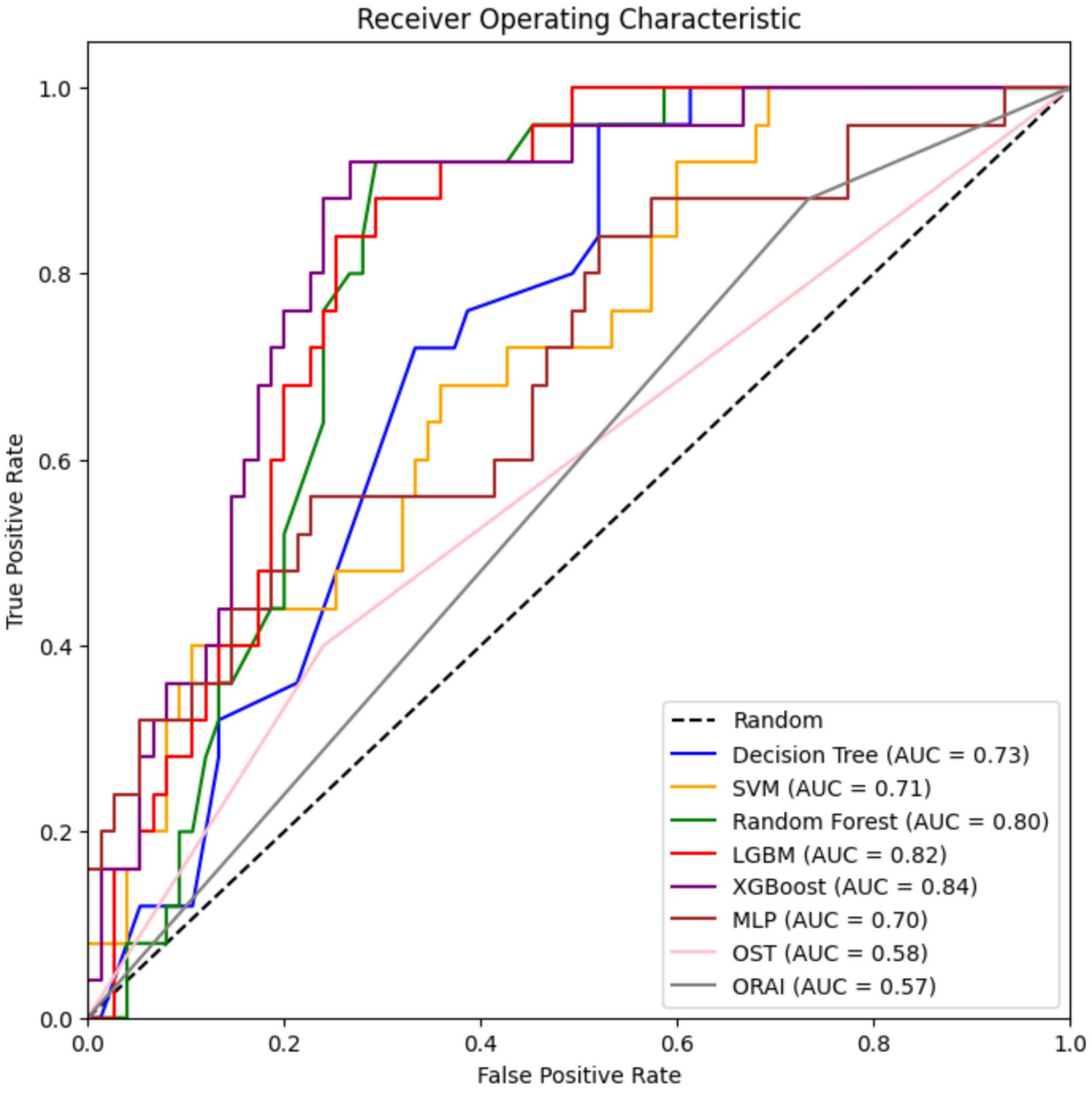

Fig. 2 The Receiver operating characteristic curve of six classification models and two traditional osteoporosis screening tools. LGBM, light gradient boosting machine; MLP, multi-layer perceptron; ORAI, osteoporosis risk assessment instrument; OST, osteoporosis self-assessment tool; SVM, support vector machine; XGBoost, extreme gradient boosting

menopause could cause an imbalance in bone formation and resorption, leading to bone loss and an increased risk of osteoporosis [18, 48].

Socio-demographic factors also play a significant role in osteoporosis risk. In our analysis, education level and income level ranked 6th and 8th in the variable importance, respectively. Consistent with our findings, a recent study showed that a higher education level was significantly associated with a heel BMD increase, while reducing the risk of osteoporosis [49]. Moreover, a higher income level was significantly associated with a femoral neck BMD increase [49]. This indicated that higher education and income levels could benefit bone health by increasing access to healthcare and contributing to healthier lifestyles [49]. Moreover, previous cross-sectional studies have shown the association between higher education level or income level and reduced risk of osteoporosis [50, 51].

In addition to variables of socio-demographic characteristics, medical history variables showed strong

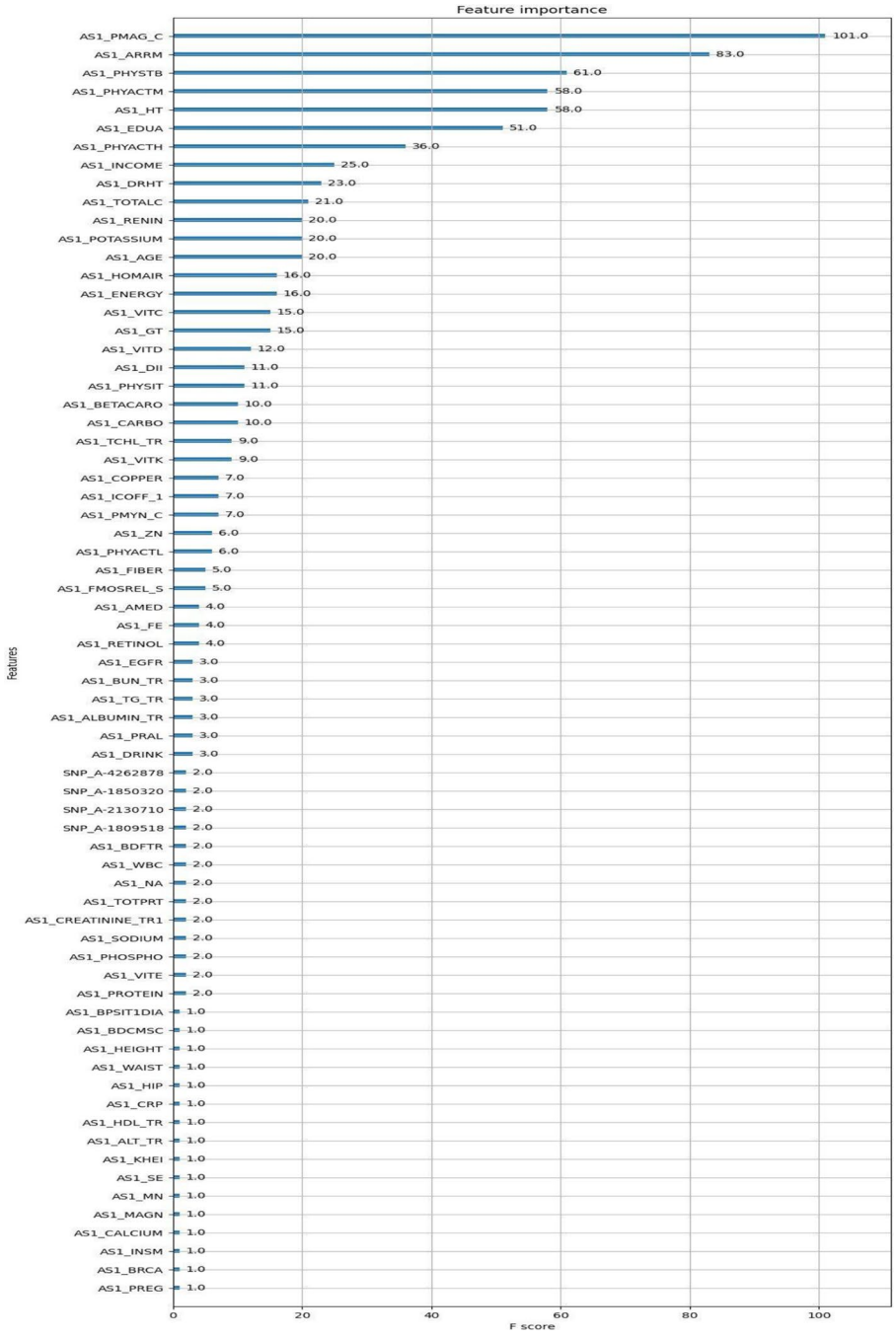

Fig. 3 (See legend on next page.)

associations with osteoporosis risk. Arthritis (osteoarthritis and rheumatoid arthritis) ranked 2nd in the variable importance. Arthritis is classified into osteoarthritis and rheumatoid arthritis. Osteoarthritis is a degenerative condition that asymmetrically affects knee and hip joints, while rheumatoid arthritis is a systemic autoimmune disease that impacts small joints, such as hands and feet [52, 53]. Arthritis was associated with inflammatory cytokines, such as interleukin-1 (IL-1), interleukin-6 (IL-6), and tumor necrosis factor-alpha (TNF-α) [54–56]. Increased levels of inflammatory cytokines could lead to elevated bone resorption, increasing the risk of osteoporosis [54–56]. Moreover, the autoimmune response of the immune system could damage bone and cartilage, increasing the risk of osteoporosis [55, 56]. In line with our findings, a cross-sectional study with 4311 subjects showed that subjects with moderate to severe osteoarthritis were associated with lower T-scores of the lumbar

(See figure on previous page.)

Fig. 3 Variable importance derived from the XGBoost technique. AS1_AGE, age; AS1_ALBUMIN_TR, albumin; AS1_ALT_TR, alanine aminotransferase; AS1_AMED, alternate Mediterranean diet score; AS1_ARRM, arthritis (osteoarthritis and rheumatoid arthritis); AS1_BDCMSC, muscle mass; AS1_BDFTR, body fat; AS1_BETACARO, beta-carotene intake; AS1_BPSIT1DIA, diastolic blood pressure; AS1_BRCA, breast cancer surgery; AS1_BUN_TR, blood urea nitrogen; AS1_CALCIUM, calcium intake; AS1_CARBO, carbohydrate intake; AS1_COPPER, copper intake; AS1_CREATININE_TR1, creatinine; AS1_CRP, C-reactive protein; AS1_DII, dietary inflammatory index; AS1_DRHT, hypertension medication; AS1_DRINK, alcohol intake status; AS1_EDUA, education level; AS1_EGFR, estimated glomerular filtration rate; AS1_ENERGY, energy intake; AS1_FE, iron intake; AS1_FIBER, fiber intake; AS1_FMOSREL_S, family history of osteoporosis (siblings); AS1_GT, gout; AS1_HDL_TR, high-density lipoprotein cholesterol; AS1_HEIGHT, height; AS1_HIP, hip circumference; AS1_HOMAIR, homeostatic model assessment for insulin resistance; AS1_HT, hypertension; AS1_ICOFF_1, frequency of coffee consumption; AS1_INCOME, income level; AS1_INSM, insomnia; AS1_KHEI, Korean healthy eating index; AS1_MAGN, magnesium intake; AS1_MN, manganese intake; AS1_NA, sodium; AS1_PHOSPHO, phosphorus intake; AS1_PHYACTH, high-intensity physical activity; AS1_PHYACTL, low-intensity physical activity; AS1_PHYACTM, moderate-intensity physical activity; AS1_PHYSIT, sedentary physical activity; AS1_PHYSTB, stable physical activity; AS1_PMAG_C, age at menopause; AS1_PMYN_C, menopausal status; AS1_POTASSIUM, potassium intake; AS1_PRAL, potential renal acid load; AS1_PREG, pregnancy experience status; AS1_PROTEIN, protein intake; AS1_RENIN, renin; AS1_RETINOL, retinol intake; AS1_SE, selenium intake; AS1_SODIUM, sodium intake; AS1_TCHL_TR, total cholesterol; AS1_TG_TR, triglyceride; AS1_TOTALC, alcohol intake; AS1_TOTPRT, total protein; AS1_VITC, vitamin C intake; AS1_VITD, vitamin D intake; AS1_VITE, vitamin E intake; AS1_VITK, vitamin K intake; AS1_WAIST, waist circumference; AS1_WBC, white blood cell; AS1_ZN, zinc intake; SNP_A-1,809,518, rs628948; SNP_A-1,850,320, rs12590815; SNP_A-2,130,710, rs238340; SNP_A-4,262,878, rs746219

spine and total hip compared with non-osteoarthritis [57]. Moreover, a cross-sectional study of 1322 Korean postmenopausal women with rheumatoid arthritis showed that 619 (46.8%) subjects were diagnosed with osteoporosis [58]. However, our study did not examine arthritis variables separately into osteoarthritis and rheumatoid arthritis. Future studies are needed to examine separately osteoarthritis and rheumatoid arthritis.

Moreover, other medical conditions were significantly associated with osteoporosis risk. Among them, hypertension emerged as a significant risk factor, with hypertension, hypertension medication, and renin levels ranking 5th, 9th, and 11th, respectively, in the variable importance. Consistent with our findings, a retrospective study showed that hypertension was significantly associated with osteoporosis risk in 2039 Chinese postmenopausal women [59]. One potential mechanism linking hypertension to osteoporosis involves the renin-angiotensin system. Renin, an enzyme secreted by kidneys contributes to the production of angiotensin II, which mediates vasoconstriction, leading to blood pressure elevation. This angiotensin II could interfere with bone formation and reduce BMD, thereby increasing the risk of osteoporosis [60, 61].

Other metabolic disorders have also been linked to osteoporosis risk. We found that gout ranked 17th in the variable importance. In line with our finding, a longitudinal study by Kwon et al. [62] showed that subjects with gout had an 11% increased risk of osteoporosis compared with subjects without gout.

Lifestyle factors were identified as key contributors to osteoporosis risk. Among them, physical activities played a crucial role, ranking 3rd, 4th, 7th, 20th, and 29th, respectively, in the variable importance. Consistent with our findings, a recent cross-sectional study found that moderate-intensity physical activity and high-intensity active physical activity could decrease the osteoporosis risk [63]. Physical activities could enhance BMD and bone strength by stimulating the bones, which could reduce the risk of osteoporosis [64, 65].

Alcohol intake was one of lifestyle factors that showed a significant association with osteoporosis. It ranked 10th in the variable importance. A meta-analysis showed that subjects who daily had 1 to 2 alcoholic drinks were likely to have a 1.34 times higher osteoporosis risk than non-alcohol drinkers [66]. Alcohol intake can inhibit osteoblast formation and stimulate osteoclast activity, which causes a decrease in bone formation and an increase in bone resorption [67, 68]. In addition, excessive alcohol consumption can elevate parathyroid hormone levels and induce oxidative stress, causing bone loss [67, 68].

Dietary intake was identified as an important determinant of osteoporosis risk. Nutrient intake variables ranked in relatively high positions in variable importance. Potassium intake ranked 12th in the variable importance. Moreover, potassium intake ranked 1st in the variable importance out of the 23 nutrient intake variables. A cross-sectional study showed that higher daily potassium intake was significantly associated with a 32% reduced risk of lumbar spine osteoporosis in 5142 postmenopausal women [69]. Dietary potassium could neutralize excess acid produced during metabolic processes, thereby maintaining the body's acid-base balance and supporting mechanisms that promote bone health [70]. Vitamin C, beta-carotene, and zinc ranked 16th, 21st, and 28th, respectively, in the variable importance. Consistently, a cross-sectional study by Kim et al. [71] found positive associations between the intakes of beta-carotene, zinc, and vitamin C and bone health in 189 postmenopausal Korean women.

Dietary quality played a significant role in osteoporosis risk. We investigated the importance of dietary quality indices for osteoporosis in women. DII ranked 19th in the variable importance, which was consistent with the findings from recent studies [72, 73]. A cross-sectional study by Li et al. [72] showed that a higher DII was significantly associated with BMD loss in the femoral neck,

intertrochanter, and total hip compared with low DII. The DII is a tool to assess the effect of an individual diet on the level of inflammation in the body [74, 75]. Consumption of pro-inflammatory foods, such as processed meats, refined grains, and high-fat dairy products, could reduce osteoblast function, activate osteoclast activity, and increase inflammation levels, leading to an increased risk of osteoporosis [74, 75].

Biochemical parameters were significantly associated with osteoporosis. Fourteen variables under the category of biochemical parameters appeared to play important roles in osteoporosis. Fourteen variables were renin, HOMA–IR, total cholesterol, eGFR, blood urea nitrogen, triglycerides, albumin, white blood cell, sodium, total protein, creatinine, CRP, high-density lipoprotein cholesterol, and alanine aminotransferase. HOMA-IR ranked 14th in the variable importance, the second highest importance after renin (11th) under the category of biochemical parameters. A prospective study by Napoli et al. [76] found that BMD increased with higher HOMA–IR in 2398 elderly adults without diabetes. This study found that CRP ranked 59th in the variable importance. A recent study by Little-Letsinger et al. [77] found a weak association between CRP and BMDs in the femoral neck and lumbar spine.

Bone turnover markers (BTMs) further contribute to understanding osteoporosis risk. In patients with bone diseases, BTMs, biomarkers found in blood and/or urine, can be used to examine the bone status [78, 79]. BTMs are classified as bone formation markers and bone resorption markers. Bone formation markers include type 1 procollagen-N-propeptide (P1NP), bone-specific alkaline phosphatase (BSAP), and osteocalcin [78, 79]. Bone resorption markers include C-terminal telopeptide of type 1 collagen (CTX), tartrate-resistant acid phosphatase 5b (TRAP 5b), and N-telopeptide of type 1 collagen (NTX) [78, 79]. In a cross-sectional study of 2327 elderly subjects aged 60 to 85 years, BSAP and NTX were inversely associated with lumbar spine BMD [80]. However, data from the Ansan and Ansung study used in this study did not provide BTMs such as CTX and P1NP.

Genetic factors also played a role in osteoporosis risk. We found 4 SNPs (rs746219, rs12590815, rs238340, and rs628948) out of the 12 SNPs ranked 41th to 44th in the variable importance. A recent study by Park et al. [14] found the association between SNPs and the risk of osteoporosis.

This study has several strengths. We utilized the Ansan and Ansung cohort study from KoGES, which is large-scale general Korean population-based cohort data, in order to construct a ML model for osteoporosis classification. We attempted to include as many osteoporosis risk factors as much as possible in the ML models. These ML models have the potential to be applicable in screening women with a high risk of osteoporosis. These ML models could be used in the field of early detection of osteoporosis, identifying risk factors, and allowing personalized osteoporosis prevention strategies. Therefore, ML models could enhance osteoporosis related-health outcomes in women, which could be beneficial in clinical and community settings.

Despite the strengths, this study has limitations. The Ansan and Ansung study of KoGES used in this study collected data using self-reported questionnaires, which could potentially have recall bias. The osteoporosis ML classification model was developed using baseline survey data from the Ansan and Ansung study of KoGES. Further studies are necessary to validate the ML classification model with follow-up data from the Ansan and Ansung study of KoGES.

Conclusions

In conclusion, we developed a ML model to classify osteoporosis in Korean women using various osteoporosis risk factors. The ML classification model using the XGBoost technique outperformed the ML classification models using the decision tree, random forest, MLP, SVM, and LGBM techniques and traditional osteoporosis screening tools using the ORAI and OST. In the variable importance using the XGBoost technique, age at menopause was the most crucial osteoporosis risk factor.

Abbreviations

AdaBoost	Adaptive boosting
aMED	Alternate Mediterranean diet score
ANN	Artificial neural network
AUC	Area under the curve
AUROC	Area under the receiver operating characteristic curve
BMD	Bone mineral density
BMI	Body mass index
BSAP	Bone-specific alkaline phosphatase
BTMs	Bone turnover markers
CAN–Pro	Computer aided nutritional analysis program
CatBoost	Categorical boosting
CRP	C-reactive protein
CTX	C-terminal telopeptide of type 1 collagen
DASH	Dietary Approaches to Stop Hypertension
DII	Dietary inflammatory index
DXA	Dual X-ray absorptiometry
eGFR	Estimated glomerular filtration rate
FPR	False positive rate
HOMA–IR	Homeostatic model assessment for insulin resistance
IL-1	Interleukin-1
IL-6	Interleukin-6
INQ	Index of nutritional quality
KHEI	Korean healthy eating index
KNHANES	Korea National Health and Nutrition Examination Survey
KNN	K-nearest neighbor
KoGES	Korean Genome and Epidemiology Study
LGBM	Light gradient boosting machine
ML	Machine learning
MLP	Multi-layer perceptron
NEAP	Net endogenous acid production
NTX	N-telopeptide of type 1 collagen
ORAI	Osteoporosis risk assessment instrument
OST	Osteoporosis self-assessment tool
P1NP	Type 1 procollagen-N-propeptide

PRAL	Potential renal acid load
QCT	Quantitative computed tomography
QUS	Quantitative ultrasound
SNP	Single nucleotide polymorphism
SQFFQ	Semi-quantitative food frequency questionnaires
SVM	Support vector machine
TNF-α	Tumor necrosis factor-alpha
TPE	Tree-structured Parzen Estimator
TPR	True positive rate
TRAP 5b	Tartrate-resistant acid phosphatase 5b
WHO	World Health Organization
XGBoost	Extreme gradient boosting

Acknowledgements
Clinical trial number: not applicable.

Author contributions
Conceptualization, Y.K.; methodology, Y.K. and S.L.; software, M.J. and S.L.; validation M.J. and S.L.; formal analysis, M.J. and S.H.; investigation, M.J. and S.H.; resources, Y.K.; data curation, Y.K., M.J., S.H., and S.L.; writing—original draft preparation, M.J.; writing—review and editing, Y.K. and S.L.; visualization, Y.K., M.J., S.H., and S.L.; supervision, Y.K.; project administration, Y.K.; funding acquisition, Y.K.; All authors have read and agreed to the published version of the manuscript.

Funding
This study was funded by the National Research Foundation of Korea (NRF), grant number NRF-2022R1F1A1063108. The NRF had no role in the study design, data analysis, or writing of this article.

Data availability
The research data that support the findings of this study have been deposited in the National Biobank of Korea (https://biobank.nih.go.kr/Desk/), the Korea Disease Control and Prevention Agency, Republic of Korea with the primary accession code NBK-2023-003.

Declarations

Ethics approval and consent to participate
This study was conducted following the guidelines of the Declaration of Helsinki, with all subjects providing written informed consent. This study was approved by the Institutional Review Board of Gyeongsang National University (GIRB-A22-NX-0073) and the Korean Health and Genomic study at the Korea National Institute of Health (NBK-2023-003).

Consent for publication
Not required.

Competing interests
The authors declare no competing interests.

Author details
[1]Department of Food and Nutrition, Gyeongsang National University, Jinju 52828, Republic of Korea
[2]Department of Computer Science and Engineering, Gyeongsang National University, Jinju 52828, Republic of Korea
[3]Department of Food and Nutrition, Institute of Agriculture and Life Science, Gyeongsang National University, Jinju 52828, Republic of Korea

Received: 16 January 2025 / Accepted: 12 March 2025
Published online: 28 March 2025

References

1. Rachner TD, Khosla S, Hofbauer LC. Osteoporosis: now and the future. Lancet. 2011;377(9773):1276–87.
2. Johnston CB, Dagar M. Osteoporosis in older adults. Med Clin North Am. 2020;104(5):873–84.
3. Moayyeri A, Warden J, Han S, Suh H, Pinedo-Villanueva R, Harvey N, et al. Estimating the economic burden of osteoporotic fractures in a multinational study: a real-world data perspective. Osteoporos Int. 2023;34(12):2121–32.
4. Föger-Samwald U, Kerschan-Schindl K, Butylina M, Pietschmann P. Age related osteoporosis: targeting cellular senescence. Int J Mol Sci. 2022;23(5).
5. Xiao P-L, Cui A-Y, Hsu C-J, Peng R, Jiang N, Xu X-H, et al. 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.
6. Service HIRA. Statistics of Diseases and Medical Practices in Life [Available from: https://www.hira.or.kr/bbsDummy.do?pgmid=HIRAA020045010000&brdScnBltNo=4&brdBltNo=2361&pageIndex=1&pageIndex2=1
7. Cosman F, de Beur SJ, LeBoff M, Lewiecki E, Tanner B, Randall S, et al. Clinician's guide to prevention and treatment of osteoporosis. Osteoporos Int. 2014;25:2359–81.
8. Pisani P, Renna MD, Conversano F, Casciaro E, Muratore M, Quarta E, et al. Screening and early diagnosis of osteoporosis through X-ray and ultrasound based techniques. World J Radiol. 2013;5(11):398.
9. Cadarette SM, Jaglal SB, Kreiger N, McIsaac WJ, Darlington GA, Tu JV. Development and validation of the osteoporosis risk assessment instrument to facilitate selection of women for bone densitometry. CMAJ. 2000;162(9):1289–94.
10. Koh L, Ben Sedrine W, Torralba T, Kung A, Fujiwara S, Chan S, et al. A simple tool to identify Asian women at increased risk of osteoporosis. Osteoporos Int. 2001;12:699–705.
11. Raisz LG. Screening for osteoporosis. N Engl J Med. 2005;353(2):164–71.
12. Bijelic R, Milicevic S, Balaban J. Risk factors for osteoporosis in postmenopausal women. Med Archives. 2017;71(1):25.
13. Shanbhogue VV, Brixen K, Hansen S. Age- and Sex-Related changes in bone microarchitecture and estimated strength: A Three-Year prospective study using HRpQCT. J Bone Min Res. 2016;31(8):1541–9.
14. Park S, Daily JW, Song MY, Kwon H-K. Gene-gene and gene-lifestyle interactions of AKAP11, KCNMA1, PUM1, SPTBN1, and EPDR1 on osteoporosis risk in middle-aged adults. Nutrition. 2020;79:110859.
15. Holm JP, Hyldstrup L, Jensen J-EB. Time trends in osteoporosis risk factor profiles: a comparative analysis of risk factors, comorbidities, and medications over twelve years. Endocrine. 2016;54:241–55.
16. Muñoz-Garach A, García-Fontana B, Muñoz-Torres M. Nutrients and dietary patterns related to osteoporosis. Nutrients. 2020;12(7):1986.
17. Lu L, Tian L. Postmenopausal osteoporosis coexisting with sarcopenia: the role and mechanisms of Estrogen. J Endocrinol. 2023;259(1).
18. Cheng C-H, Chen L-R, Chen K-H. Osteoporosis due to hormone imbalance: an overview of the effects of Estrogen deficiency and glucocorticoid overuse on bone turnover. Int J Mol Sci. 2022;23(3):1376.
19. Eastell R, O'Neill TW, Hofbauer LC, Langdahl B, Reid IR, Gold DT, et al. Postmenopausal osteoporosis. Nat Reviews Disease Primers. 2016;2(1):1–16.
20. Black JE, Kueper JK, Williamson TS. An introduction to machine learning for classification and prediction. Fam Pract. 2023;40(1):200–4.
21. Zhao H, Zhang X, Xu Y, Gao L, Ma Z, Sun Y, et al. Predicting the risk of hypertension based on several easy-to-collect risk factors: a machine learning method. Front Public Health. 2021;9:619429.
22. Kim H, Hwang S, Lee S, Kim Y. Classification and prediction on hypertension with blood pressure determinants in a deep learning algorithm. Int J Environ Res Public Health. 2022;19(22):15301.
23. Gutiérrez-Esparza G, Pulido T, Martínez-García M, Ramírez-delReal T, Groves-Miralrio LE, Márquez-Murillo MF, et al. A machine learning approach to personalized predictors of dyslipidemia: a cohort study. Front Public Health. 2023;11:1213926.
24. Ravaut M, Harish V, Sadeghi H, Leung KK, Volkovs M, Kornas K, et al. Development and validation of a machine learning model using administrative health data to predict onset of type 2 diabetes. JAMA Netw Open. 2021;4(5):e2111315–e.
25. Ming C, Viassolo V, Probst-Hensch N, Chappuis PO, Dinov ID, Katapodi MC. Machine learning techniques for personalized breast cancer risk prediction: comparison with the BCRAT and BOADICEA models. Breast Cancer Res. 2019;21:1–11.
26. Inui A, Nishimoto H, Mifune Y, Yoshikawa T, Shinohara I, Furukawa T, et al. Screening for osteoporosis from blood test data in elderly women using a machine learning approach. Bioengineering. 2023;10(3):277.
27. Bui HM, Ha MH, Pham HG, Dao TP, Nguyen T-TT, Nguyen ML, et al. Predicting the risk of osteoporosis in older Vietnamese women using machine learning approaches. Sci Rep. 2022;12(1):20160.

28. Ou Yang W-Y, Lai C-C, Tsou M-T, Hwang L-C. Development of machine learning models for prediction of osteoporosis from clinical health examination data. Int J Environ Res Public Health. 2021;18(14):7635.
29. Kwon Y, Lee J, Park JH, Kim YM, Kim SH, Won YJ, et al. editors. Osteoporosis pre-screening using ensemble machine learning in postmenopausal Korean women. Healthcare: MDPI; 2022.
30. Shim J-G, Kim DW, Ryu K-H, Cho E-A, Ahn J-H, Kim J-I, et al. Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women. Archives Osteoporos. 2020;15:1–9.
31. Kim Y, Han B-G, Group K. Cohort profile: the Korean genome and epidemiology study (KoGES) consortium. Int J Epidemiol. 2017;46(2):e20–e.
32. Prevention NIoHKCfDCa. Examination and Survey Quality Control of the Korean Genome and Epidemiology Study [Available from: https://www.kdca.go.kr/contents.es?mid=a40504010000
33. Sözen T, Özışık L, Başaran NÇ. An overview and management of osteoporosis. Eur J Rheumatol. 2017;4(1):46.
34. Lee K. Association of osteosarcopenic obesity and its components: osteoporosis, sarcopenia and obesity with insulin resistance. J Bone Miner Metab. 2020;38:695–701.
35. Lee J, Oh K-H, Park S-K. Dietary micronutrients and risk of chronic kidney disease: a cohort study with 12 year follow-up. Nutrients. 2021;13(5):1517.
36. Vahid F, Hatami M, Sadeghi M, Ameri F, Faghfoori Z, Davoodi SH. The association between the index of nutritional quality (INQ) and breast cancer and the evaluation of nutrient intake of breast cancer patients: A case-control study. Nutrition. 2018;45:11–6.
37. Storz MA, Ronco AL. Reduced dietary acid load in US vegetarian adults: results from the National health and nutrition examination survey. Food Sci Nutr. 2022;10(6):2091–100.
38. Fung TT, McCullough ML, Newby P, Manson JE, Meigs JB, Rifai N, et al. Diet-quality scores and plasma concentrations of markers of inflammation and endothelial dysfunction. Am J Clin Nutr. 2005;82(1):163–73.
39. Jennings A, Mulligan AA, Khaw K-T, Luben RN, Welch AA. A mediterranean diet is positively associated with bone and muscle health in a non-Mediterranean region in 25,450 men and women from EPIC-Norfolk. Nutrients. 2020;12(4):1154.
40. Fung TT, Chiuve SE, McCullough ML, Rexrode KM, Logroscino G, Hu FB. Adherence to a DASH-style diet and risk of coronary heart disease and stroke in women. Arch Intern Med. 2008;168(7):713–20.
41. Du S, Chen J, Kim H, Walker ME, Lichtenstein AH, Chatterjee N, et al. Plasma protein biomarkers of healthy dietary patterns: results from the atherosclerosis risk in communities study and the Framingham heart study. J Nutr. 2023;153(1):34–46.
42. Shivappa N, Steck SE, Hurley TG, Hussey JR, Hébert JR. Designing and developing a literature-derived, population-based dietary inflammatory index. Public Health Nutr. 2014;17(8):1689–96.
43. Yook S-M, Park S, Moon H-K, Kim K, SHIM J-E, Hwang J-Y. Development of Korean healthy eating index for adults using the Korea National health and nutrition examination survey data. J Nutr Health. 2015:419–28.
44. Dorogush AV, Ershov V, Gulin A. CatBoost: gradient boosting with categorical features support. ArXiv Preprint arXiv:181011363. 2018.
45. Akiba T, Sano S, Yanase T, Ohta T, Koyama M, editors. Optuna: A next-generation hyperparameter optimization framework. Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining; 2019.
46. Jones A, Enticott J, Ebeling P, Mishra G, Teede H, Vincent A. Bone health in women with premature ovarian insufficiency/early menopause: a 23-year longitudinal analysis. Hum Reprod. 2024;39(5):1013–22.
47. He Y, Huang J, Jiang G, Wang H, Zhao J, Chen Z, et al. Menarche age exceed 17 years and menopausal age smaller than 48 years May affect prevalence of osteoporosis for Chinese women. Archives Osteoporos. 2021;16(1):123.
48. de Villiers TJ. Bone health and menopause: osteoporosis prevention and treatment. Best Pract Res Clin Endocrinol Metab. 2024;38(1):101782.
49. Duan J-Y, You R-X, Zhou Y, Xu F, Lin X, Shan S-K et al. Assessment of causal association between the socio-economic status and osteoporosis and fractures: a bidirectional Mendelian randomization study in European population. J Bone Miner Res. 2024:zjae060.
50. Heidari B, Hosseini R, Javadian Y, Bijani A, Sateri MH, Nouroddini HG. Factors affecting bone mineral density in postmenopausal women. Archives Osteoporos. 2015;10:1–7.
51. Du Y, Zhao L-J, Xu Q, Wu K, Deng H-W. Socioeconomic status and bone mineral density in adults by race/ethnicity and gender: the Louisiana osteoporosis study. Osteoporos Int. 2017;28:1699–709.
52. Park J, Mendy A, Vieira ER. Various types of arthritis in the united States: prevalence and age-related trends from 1999 to 2014. Am J Public Health. 2018;108(2):256–8.
53. Mohammed A, Alshamarri T, Adeyeye T, Lazariu V, McNutt L-A, Carpenter DO. A comparison of risk factors for osteo-and rheumatoid arthritis using NHANES data. Prev Med Rep. 2020;20:101242.
54. Huang K, Cai H. The interplay between osteoarthritis and osteoporosis: mechanisms, implications, and treatment considerations–A narrative review. Exp Gerontol. 2024;197:112614.
55. Deng Y, Wong MCS. Association between rheumatoid arthritis and osteoporosis in Japanese populations: a Mendelian randomization study. Arthritis Rheumatol. 2023;75(8):1334–43.
56. Llorente I, García-Castañeda N, Valero C, González-Álvaro I, Castañeda S. Osteoporosis in rheumatoid arthritis: dangerous liaisons. Front Med. 2020;7:601618.
57. Choi E-S, Shin HD, Sim JA, Na YG, Choi W-J, Shin D-D, et al. Relationship of bone mineral density and knee osteoarthritis (Kellgren-Lawrence grade): fifth Korea National health and nutrition examination survey. Clin Orthop Surg. 2021;13(1):60.
58. Lee J-H, Sung Y-K, Choi C-B, Cho S-K, Bang S-Y, Choe J-Y, et al. The frequency of and risk factors for osteoporosis in Korean patients with rheumatoid arthritis. BMC Musculoskelet Disord. 2016;17:1–7.
59. Chai H, Ge J, Li L, Li J, Ye Y. Hypertension is associated with osteoporosis: a case-control study in Chinese postmenopausal women. BMC Musculoskelet Disord. 2021;22:1–7.
60. Azeez TA. Osteoporosis and cardiovascular disease: a review. Mol Biol Rep. 2023;50(2):1753–63.
61. Tamargo J, Caballero R, Delpón E. The renin–angiotensin system and bone. Clin Rev Bone Miner Metab. 2015;13:125–48.
62. Kwon MJ, Park JY, Kim SG, Kim J-K, Lim H, Kim J-H, et al. Potential association of osteoporosis and not osteoporotic fractures in patients with gout: A longitudinal follow-up study. Nutrients. 2022;15(1):134.
63. Park H-Y, Jung W-S, Kim S-W, Lim K. Relationship between sarcopenia, obesity, osteoporosis, and cardiometabolic health conditions and physical activity levels in Korean older adults. Front Physiol. 2021;12:706259.
64. Tong X, Chen X, Zhang S, Huang M, Shen X, Xu J, et al. The effect of exercise on the prevention of osteoporosis and bone angiogenesis. Biomed Res Int. 2019;2019(1):8171897.
65. Yuan Y, Chen X, Zhang L, Wu J, Guo J, Zou D, et al. The roles of exercise in bone remodeling and in prevention and treatment of osteoporosis. Prog Biophys Mol Biol. 2016;122(2):122–30.
66. Cheraghi Z, Doosti-Irani A, Almasi-Hashiani A, Baigi V, Mansournia N, Etminan M, et al. The effect of alcohol on osteoporosis: A systematic review and meta-analysis. Drug Alcohol Depend. 2019;197:197–202.
67. Cho Y, Choi S, Kim K, Lee G, Park SM. Association between alcohol consumption and bone mineral density in elderly Korean men and women. Archives Osteoporos. 2018;13:1–8.
68. Zhu K, Prince RL. Lifestyle and osteoporosis. Curr Osteoporos Rep. 2015;13:52–9.
69. Ha J, Kim S-A, Lim K, Shin S. The association of potassium intake with bone mineral density and the prevalence of osteoporosis among older Korean adults. Nutr Res Pract. 2020;14(1):55.
70. Singh W, Kushwaha P, Potassium. A frontier in osteoporosis. Hormone and Metabolic Research; 2024.
71. Kim DE, Cho SH, Park HM, Chang YK. Relationship between bone mineral density and dietary intake of β-carotene, vitamin C, zinc and vegetables in postmenopausal Korean women: a cross-sectional study. J Int Med Res. 2016;44(5):1103–14.
72. Li S, Zeng M. The association between dietary inflammation index and bone mineral density: results from the united States National health and nutrition examination surveys. Ren Fail. 2023;45(1):2209200.
73. Zhao S, Gao W, Li J, Sun M, Fang J, Tong L, et al. Dietary inflammatory index and osteoporosis: the National health and nutrition examination survey, 2017–2018. Endocrine. 2022;78(3):587–96.
74. Fang Y, Zhu J, Fan J, Sun L, Cai S, Fan C, et al. Dietary inflammatory index in relation to bone mineral density, osteoporosis risk and fracture risk: a systematic review and meta-analysis. Osteoporos Int. 2021;32:633–43.
75. Song D, Kim J, Kang M, Park J, Lee H, Kim D-Y, et al. Association between the dietary inflammatory index and bone markers in postmenopausal women. PLoS ONE. 2022;17(3):e0265630.

76. Napoli N, Conte C, Pedone C, Strotmeyer ES, Barbour KE, Black DM, et al. Effect of insulin resistance on BMD and fracture risk in older adults. J Clin Endocrinol Metabolism. 2019;104(8):3303–10.
77. Little-Letsinger SE. Serum high sensitivity C-reactive protein poorly predicts bone mineral density: A NHANES 2017–2020 analysis. PLoS ONE. 2023;18(10):e0288212.
78. Schini M, Vilaca T, Gossiel F, Salam S, Eastell R. Bone turnover markers: basic biology to clinical applications. Endocr Rev. 2023;44(3):417–73.
79. Vasikaran S, Thambiah SC, Tan RZ, Loh TP. The use of Bone-Turnover markers in Asia-Pacific populations. Ann Lab Med. 2024;44(2):126–34.
80. Zhu Z, Zhou H, Wang Y, Yao X. Associations between bone turnover markers and bone mineral density in older adults. J Orthop Surg. 2021;29(1):2309499020987653.

Publisher's note