Bone Reports 24 (2025) 101826

Contents lists available at ScienceDirect

Bone Reports

journal homepage: www.elsevier.com/locate/bonr

A simple and user-friendly machine learning model to detect osteoporosis in health examination populations in Southern Taiwan

Wei-Chin Huang [a], I-Shu Chen [a], Hsien-Chung Yu [a], Chi-Shen Chen [a,c], Fu-Zong Wu [b], Chiao-Lin Hsu [a,*], Pin-Chieh Wu [a,c,*]

[a] Health Management Center, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan
[b] Department of Radiology, Kaohsiung Veterans General Hospital, Kaohsiung, Taiwan
[c] Department of Pharmacy, Chia Nan University of Pharmacy & Science, Tainan, Taiwan

ARTICLE INFO

Keywords:
Osteoporosis
Machine learning
Dual-energy X-ray absorptiometry
Artificial neural networks
Artificial intelligence

ABSTRACT

Background: Osteoporosis is a growing public health concern in aging populations such as Taiwan, where limited utilization of dual-energy X-ray absorptiometry (DXA) often leads to underdiagnosis and even delayed treatment. Therefore, we leveraged machine learning (ML) and aimed to develop a simple and easily accessible model that effectively identifies individuals at high risk of osteoporosis.

Methods: This retrospective analysis enrolled 5510 men aged ≥50 years and 4720 postmenopausal women who underwent DXA at the Kaohsiung Veterans General Hospital, with another cohort of 610 men and 523 women for validation. We developed separate models for men and women using decision trees, random forests, support vector machines, k-nearest neighbors, extreme gradient boosting, and artificial neural networks (ANNs) to predict osteoporosis. Furthermore, we compared each model with the traditional Osteoporosis Self-Assessment Tool for Asians (OSTA) model.

Results: We identified age, height, weight, and BMI as variables for our prediction model and evaluated the model's performance using the area under the receiver operating characteristic curve (AUC). The ANN model significantly outperformed the OSTA model and all the other ML models for both men and women (AUC: 0.67 for men; 0.77 for women). The validation data for the ANN model showed similar AUCs for both men and women.

Conclusion: This study developed ML models to help identify individuals at high risk of osteoporosis in postmenopausal women and men aged ≥50 years in southern Taiwan. Our ML models, especially the ANN model, surpassed the OSTA model and consistently performed well across different populations.

1. Introduction

Osteoporosis, marked by declining bone mineral density (BMD) and deteriorating microarchitecture, is a growing public health concern in aging populations like Taiwan (Consensus Development Conference, 1993). In 2013, the prevalence among individuals aged ≥80 years was significantly higher than those aged 50–59 years, increasing from 2217 to 16,580 per 100,000 in females and from 724 to 6227 per 100,000 in males, mirroring the rise in major osteoporotic fractures. This highlights the concerning impact of population aging on skeletal health in Taiwan (Wang et al., 2017). Limited dual-energy X-ray absorptiometry (DXA) screening due to subtle symptoms leads to underdiagnosis and treatment (Yang et al., 2006). This challenge is based on the current diagnostic limitations and the lack of existing predictive tools. Therefore, a refined approach to effectively screen and detect individuals at greatest risk, leveraging our understanding of osteoporosis risk factors, is required.

Factors associated with osteoporosis include age, sex, height, weight, body mass index (BMI), body fat percentage, waist circumference (WC), waist-hip ratio (WHR), diet, smoking, alcohol consumption, physical activities, and co-morbidities such as hypertension (HTN), diabetes mellitus (DM), hyperlipidemia, atherosclerotic cardiovascular disease (ASCVD), thyroid disease, chronic liver disease, and autoimmune diseases (Chai et al., 2021; Delitala et al., 2020; Drinkwater, 1994; Ensrud and Crandall, 2017; Godos et al., 2022; Goldring and Gravallese, 2000; Krall and Dawson-Hughes, 1999; Pirih et al., 2012; Tao et al., 2023; Yaturu, 2009). Based on these risk factors, studies have developed tools for predicting the risk of osteoporosis, such as Simple Calculated Osteoporosis Risk Estimation-Reported, Osteoporosis Pre-screening Risk

* Corresponding authors at: Health Management Center, Kaohsiung Veterans General Hospital, 386, Ta-Chung 1st Road, Kaohsiung 813, Taiwan.
E-mail addresses: clhsu@vghks.gov.tw (C.-L. Hsu), wpchieh@vghks.gov.tw (P.-C. Wu).

https://doi.org/10.1016/j.bonr.2025.101826
Received 30 July 2024; Received in revised form 25 November 2024; Accepted 10 January 2025
Available online 11 January 2025

Assessment, and Osteoporosis Index of Risk. However, these were either overly complex – requiring extensive variable analysis, making them challenging to apply – or too simplistic, such as the Osteoporosis Self-Assessment Tool for Asians (OSTA), resulting in lower predictive accuracy (Yang et al., 2023).

Machine learning (ML) uses statistics to classify data by analyzing patterns and features, enhancing its ability to predict and categorize diverse data types intelligently. Shim et al. (2020) developed an ML model to predict osteoporosis only in postmenopausal women. Wu and Park (2023) developed a model for both men and women; however, their BMDs were obtained from the tibia or radius using quantitative ultrasound (QUS). Ou Yang et al. (2021) formulated a model for both sexes using DXA measurement; however, the osteoporosis prevalence was notably lower than that of the local population, and this may not accurately represent the authentic demographic profile. These studies have shown limitations, including narrow population focus, varied BMD measurement methods, and discrepancies in the reported osteoporosis prevalence, compared with actual demographic profiles. These constraints highlight the need for a more comprehensive and accurate model that reflects the osteoporosis risk across a broader Taiwanese population.

Therefore, we aimed to develop a ML model that effectively identifies individuals at high risk of osteoporosis, improving upon the narrow focus and varied methodologies of previous studies.

2. Patients and methods

2.1. Data source and study population

This study utilized data retrospectively extracted from the medical records of a health management center at a medical center in Southern Taiwan. We included people presenting for health check-ups aged ≥20 years who underwent DXA of the lumbar spine and bilateral hip joints at the health management center in Kaohsiung Veterans General Hospital between January 2016 and May 2023. The study involved 24,780 participants, and we excluded premenopausal women and men aged <50 years. The population was divided into two groups: one for model training between January 2016 and December 2022 and another for model validation between January 2023 and May 2023. The model training and validation groups comprised 5510 men and 4272 women and 610 men and 523 women, respectively (Fig. 1). The Institutional Review Board of the Kaohsiung Veterans General Hospital reviewed and approved this study on July 17th, 2023 (KSVGH23-CT8-08), which used anonymized data retrospectively, and the need for informed consent was waived. It was conducted according to the principles of the Declaration of Helsinki.

2.2. Data collection and BMD measurements

The detailed information on demographic characteristics, such as age, education level, and personal lifestyle, from the database was based on past interviews by physicians in person. Trained investigators used

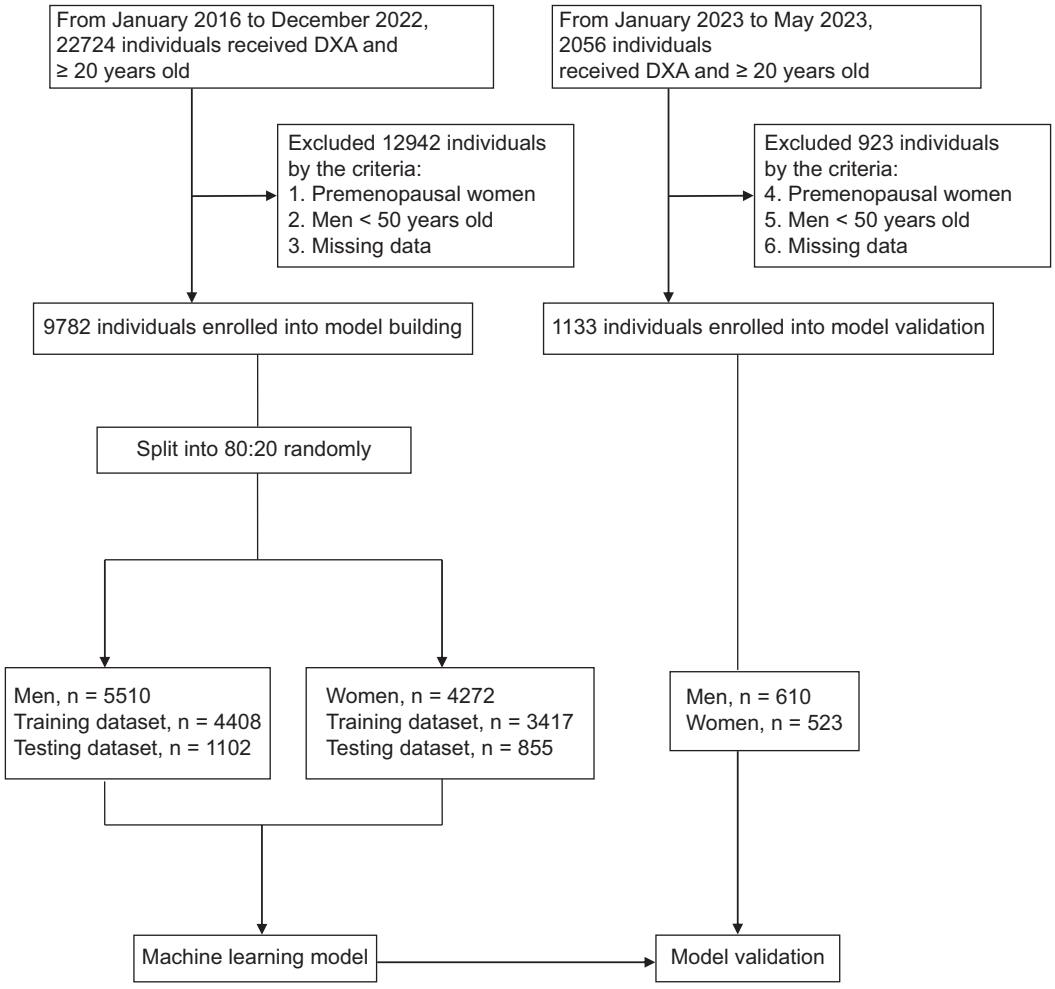

Fig. 1. Flowchart of inclusion, exclusion, and processing of model training.
DXA: dual-energy x-ray absorptiometry.

calibrated instruments and followed the manufacturers' instructions to conduct all physical measurements. After this verification process, the data were anonymized and subsequently entered into the hospital's database for researchers to utilize. Data collected on co-morbidities included data regarding HTN, DM, hyperlipidemia, ASCVD, thyroid disease, chronic liver disease, and rheumatoid arthritis (RA). Thyroid disease was defined as hyper- or hypothyroidism, and liver disease was defined as hepatitis B or C and liver cirrhosis. Disorders such as metabolic-associated fatty liver disease, alcohol-associated liver disease, and biliary disorders cannot be directly confirmed through patient interviews and require further imaging examinations, thus, these conditions were excluded as variables in the analysis. Education level was divided into three categories: primary school or below, junior or senior high school, and college or above. Diet was classified as vegetarian and non-vegetarian. Smoking status was classified as heavy ($\geq$20 pack-years), light, or non-smoker ($<$20 pack-years). Alcohol consumption was categorized as high-frequency (a person who consumed $\geq$3 times in 1 week), low-frequency, or non-drinker ($<$3 times in 1 week). Physical activity was classified based on exercise time ($\geq$150 min and $<$ 150 min weekly). All participants underwent DXA (iDXA, GE-Lunar, GE Healthcare, IL, USA) evaluation, targeting the L1–L4 segments of the lumbar spine and the entire hip region. Osteoporosis was defined as a T-score ≤ -2.5, derived from USA combined National Health and Nutrition Examination Survey and the Lunar femur and AP spine reference database, based on the World Health Organization definition (Ensrud and Crandall, 2017).

2.3. ML model development

The ML models employed in this study included decision trees (DT), random forests (RF), support vector machines (SVM), k-nearest neighbors (KNN), extreme gradient boosting (XGB), and artificial neural networks (ANN). Libraries from Scikit-learn (version 1.2.2) were used for DT, RF, and SVM algorithms. XGB (version 2.0.3) was implemented using its dedicated package, whereas ANN was developed using an Application Programming Interface from PyTorch (version 2.1.0 + cu121). The training and testing data in each training section of the DT, RF, SVM, XGB, KNN, and ANN models were randomly divided in an 8:2 ratio and internally validated using the k-fold cross validation.

Participants were stratified into two cohorts and training was conducted separately for men and women, based on variables that potentially influence BMD, comprising age; height; weight; BMI; body fat percentage; WC; WHR; diet; alcohol consumption; smoking status; physical activity; history of HTN, DM, hyperlipidemia, ASCVD, thyroid disease, chronic liver disease, and autoimmune disease; steroid usage; and hormone replacement therapy for women. Laboratory data included fasting glucose, glycohemoglobin, total cholesterol, high- and low-density lipoprotein cholesterol, triglycerides, uric acid, glutamate oxaloacetate transaminase, glutamate pyruvate transaminase, alkaline phosphatase, albumin, blood urea nitrogen, serum creatinine, and estimated glomerular filtration rate measurements. The glomerular filtration rate was calculated by the Modification of Diet in Renal Disease formula, as Eq. (1) (see Appendix A).

Regarding the hyperparameters, default values were set for DT, RF, and KNN algorithms within their respective libraries. The SVM's maximum iteration was capped at 100 to avoid prolonged computation times. XGB was configured with 100 estimators and a learning rate of 0.3. The ANN comprised 10 layers, including input and output layers, with a 0.2 dropout rate in each layer to prevent overfitting. We monitored the training and validation dataset in each epoch to ensure early stopping. The output layer's activation function was softmax, translating the output into class probabilities, whereas the other layers used the leaky ReLU. The learning rate for ANN was 0.0009, with a batch size of 10 and 50 epochs. The ANN model also incorporated class weight to address data imbalance. The optimal setting of the hyperparameters was obtained using k-fold cross-validation. Each model's final selected hyperparameters are presented in Supplementary Table 1. We applied the testing data to these ML models and the OSTA scores to evaluate their performance. The OSTA score was calculated as 0.2 [weight(kg) − age(year)], and a cutoff value of <-4 was considered to indicate osteoporosis (Koh et al., 2001).

2.4. Statistical analysis

The demographic data for continuous data were presented as means (standard deviation), and categorical data were presented as numbers (percentages). Statistical analysis was conducted using IBM SPSS statistics (version 25.0, IBM, NY, USA). The significance of the differences between the variables was determined using a *t*-test for continuous variables and Chi-squared tests for categorical variables. Statistical significance was set at $p < 0.05$.

The specificity, sensitivity, accuracy, and area under the receiver operating characteristic (AUC) values with a 95 % confidence interval (CI) were calculated and compared. Cutoff points were selected by identifying points on the receiver operating characteristic (ROC) curve that were the closest to the upper-left corner.

3. Results

Notably, 9782 participants were enrolled in model training in this study, including 5510 men and 4272 women. According to the DXA results, 801 (14.5 %) men and 1388 (32.5 %) women had osteoporosis. Men and women with osteoporosis were older and had lower height, weight, BMI, body fat percentage, and WC, compared with the corresponding groups without osteoporosis. Additionally, there was a higher proportion of vegetarians and lower proportions of individuals who exercised >150 min per week or consumed alcohol more than three times per week in the osteoporosis group than in the group without osteoporosis. An increased prevalence of ASCVD and DM was observed in men with osteoporosis than in those without osteoporosis. Women with osteoporosis had a lower WHR, education level, and less frequent use of HRT but had higher prevalence of chronic liver disease and lower prevalence of HTN than did those without osteoporosis (Table 1).

We used six ML methods (DT, RF, SVM, KNN, XGB, and ANN) to analyze data in our research. We split the dataset into training and testing groups and inputted these variables into analysis for men and women, with respective models, to improve the reliability of our results. Both male and female groups showed no significant difference in patient characteristics, co-morbidities, or laboratory data between the training and testing datasets, except for HTN history in men and heavy smoking in women. Specifically, within the male group, the training group exhibited a higher rate of HTN, compared with the testing group. However, within the female group, the training group had fewer heavy smokers, compared with the testing group (Table 2).

Notably, several demographic variables were associated with BMD according to the univariate analysis; however, there was no adjustment for other variables. Therefore, we applied permutation feature importance (PFI) to select features that enhance the performance of our ML model. We used PFI to rank features based on their importance scores and selected the most significant subset for further analysis. The models were repeatedly retrained to determine the optimal set of features that maximized effectiveness. Finally, we identified four key variables (age, height, weight, and BMI) for the osteoporosis prediction model.

For the performance of osteoporosis prediction among men, the models' AUC values ranged from 0.51 to 0.67, with the ANN model exhibiting the highest AUC value of 0.67 (95 % CI: 0.622–0.715), indicating its superior predictive ability. The SVM and XGB models followed, with AUC values of 0.55 (95 % CI: 0.506–0.601) and 0.53 (95 % CI: 0.482–0.575), respectively. Despite having the highest specificity of 0.99 and accuracy of 0.85, the OSTA model demonstrated the lowest AUC value of 0.51 (95 % CI: 0.493–0.535; Table 3). Similarly, for the performance of osteoporosis prediction among women, the AUC values

Table 1
Comparison of characteristics of the patients in the non-osteoporosis and osteoporosis cohorts.

Characteristics of participants	Men		Women	
Variables	Non-osteoporosis (n = 4709)	Osteoporosis (n = 801)	Non-osteoporosis (n = 2884)	Osteoporosis (n = 1388)
Age, years (mean [SD])	60.3 (6.9)	62.2 (7.5)***	59.2 (6.8)	63.27 (7.0)***
Education level, %				
Junior high school or below	224 (4.8)	51 (6.4)	279 (9.7)	264 (19.0)*
Senior high school	1329 (28.2)	241 (30.1)	1113 (38.6)	533 (38.4)
College or above	3156 (67.0)	509 (63.5)	1492 (51.7)	591 (42.6)
Height, cm (mean [SD])	170.0 (5.6)	168.0 (5.5)***	158.4 (5.1)	156.2 (5.3)***
Weight, kg (mean [SD])	72.3 (9.0)	67.0 (8.7)***	58.3 (7.6)	53.7 (6.9)***
BMI, kg/m^2 (mean [SD])	25.0 (2.7)	23.7 (2.8)***	23.3 (2.9)	22.0 (2.8)***
Body fat percentage (mean [SD])	22.1 (5.3)	21.4 (5.8)**	28.7 (5.7)	27.4 (5.7)***
Waist circumference, cm (mean [SD])	88.8 (7.6)	86.4 (7.7)***	81.3 (7.8)	78.7 (7.9)***
Waist hip ratio (mean [SD])	0.9 (0.1)	0.9 (0.1)	0.82 (0.1)	0.81 (0.1)*
Diet, %				
Non-vegetarian	4540 (96.4)	756 (94.4)*	2735 (94.8)	1281 (92.3)**
Vegetarian	169 (3.6)	45 (5.6)	149 (5.2)	107 (7.7)
Weekly alcohol consumption, %				
<3 times/week	4105 (87.2)	721 (90.0)*	2796 (96.9)	1365 (98.3)**
≥3 times/week	604 (12.8)	80 (10.0)	88 (3.1)	23 (1.7)
Smoked 20 pack-years, %				
<20 pack-years	3247 (69.0)	530 (66.2)	2832 (98.2)	1366 (98.4)
≥20 pack-years	1462 (31.0)	271 (33.8)	52 (1.8)	22 (1.6)
At least 150 min of weekly exercise, %				
No	2509 (53.3)	468 (58.4)**	1808 (62.7)	826 (59.5)*
Yes	2200 (46.7)	333 (41.6)	1076 (37.3)	562 (40.5)
Hypertension, %				
No	3028 (64.3)	511 (63.8)	595 (20.6)	1051 (75.7)*
Yes	1681 (35.7)	290 (36.2)	2289 (79.4)	337 (24.3)
Diabetes, %				
No	781 (16.6)	109 (13.6)*	2569 (89.1)	1248 (89.9)
Yes	3928 (83.4)	692 (86.4)	315 (10.9)	140 (10.1)
Hyperlipidemia, %				
No	3222 (68.4)	522 (65.2)	2091 (72.5)	1002 (72.2)
Yes	1487 (31.6)	249 (34.8)	793 (27.5)	386 (27.8)
ASCVD, %				
No	4266 (90.6)	693 (86.5)**	2810 (97.4)	1338 (96.4)
Yes	443 (9.4)	108 (13.5)	74 (2.6)	50 (3.6)
Autoimmune disease, %				
No	4632 (98.4)	790 (98.6)	2796 (96.9)	1342 (96.7)
Yes	77 (1.6)	11 (1.4)	88 (3.1)	46 (3.3)
Thyroid disease, %				
No	4512 (95.8)	766 (95.6)	2753 (95.5)	1331 (95.9)
Yes	197 (4.2)	35 (4.4)	131 (4.5)	57 (4.1)
Chronic liver disease, %				
No	3975 (84.4)	674 (84.1)	2579 (89.4)	1192 (85.9)**
Yes	734 (15.6)	127 (15.9)	305 (10.6)	196 (14.1)
Hormone therapy, %				
No	X	X	2805 (97.3)	1362 (98.1)*
Yes	X	X	79 (2.7)	26 (1.9)
Albumin, g/dL	4.4 (0.2)	4.4 (0.2)	4.4 (0.3)	4.4 (0.3)
GOT, U/L	23.4 (13.3)	23.4 (9.8)	23.4 (14.5)	22.8 (9.6)
GPT, U/L	27.4 (23.7)	27.6 (20.4)	27.7 (26.3)	26.7 (18.9)
ALK-P, U/L	62.1 (12.8)	62.7 (13.6)	61.9 (12.2)	61.5 (10.7)
GGT, U/L	29.6 (30.4)	31.0 (35.1)	29.3 (24.0)	28.5 (21.8)
Total bilirubin, mg/dL	0.8 (0.4)	0.8 (0.3)	0.81 (0.4)	0.82 (0.3)
BUN, mg/dL	12.9 (3.7)	12.9 (3.7)	12.9 (4.2)	12.7 (3.7)
Creatinine, mg/dL	0.9 (0.2)	0.9 (0.2)	0.9 (0.3)	0.9 (0.2)
eGFR, mL/min/1.73 m^2	85.9 (16.1)	86.0 (17.0)	85.3 (16.2)	86.1 (16.3)
Uric acid, mg/dL	5.7 (1.4)	5.7 (1.4)	5.7 (1.4)	5.7 (1.4)
Fasting sugar, mg/dL	94.2 (20.2)	95.3 (21.1)	94.8 (19.8)	94.7 (20.5)
HbA1c, %	5.8 (0.7)	5.8 (0.7)	5.8 (0.7)	5.8 (0.7)
Total cholesterol, mg/dL	195.8 (36.2)	196.1 (40.0)	195.3 (8.0)	195.1 (35.5)
HDL, mg/dL	53.4 (13.7)	53.4 (14.6)	53.1 (13.5)	53.4 (13.4)
LDL, mg/dL	118.5 (31.8)	118.2 (35)	116.8 (31.2)	117.3 (31.3)
TG, mg/dL	122.4 (88.6)	127.1 (140.4)	123.4 (101.1)	121.8 (79.3)

Values are presented as numbers (%) or mean (standard deviation [SD]).
*Significant difference in Chi-square test and t-test at $P < 0.05$, **$P < 0.01$, and ***$P < 0.001$.
HbA1c: glycohemoglobin; HDL: high-density lipoprotein cholesterol; LDL: low-density lipoprotein cholesterol; TG: triglycerides; GOT: glutamate oxaloacetate transaminase; GPT: glutamate pyruvate transaminase; ALK-P: alkaline phosphatase; BUN: blood urea nitrogen; eGFR: estimated glomerular filtration rate; GGT: gamma-glutamyl transferase; ASCVD: atherosclerotic cardiovascular disease, BMI: body mass index.

Table 2
Comparison of characteristics of the patients in the training and testing datasets.

Variables	Men		Women	
	Training (n = 4408)	Testing (n = 1102)	Training (n = 3417)	Testing (n = 855)
Age, years (mean [SD])	60.5 (7.0)	60.7 (7.1)	60.5 (7.1)	60.6 (7.4)
Education level, %				
Junior high school or below	220 (5.0)	55 (5.0)	428 (12.5)	115 (13.5)
Senior high school	1246 (28.3)	324 (29.4)	1307 (38.2)	339 (39.6)
College or above	2942 (66.7)	723 (65.6)	1682 (49.2)	401 (46.9)
Height, cm (mean [SD])	169.8 (5.7)	167.7 (5.7)	157.6 (5.3)	157.8 (5.3)
Weight, kg (mean [SD])	71.5 (9.1)	71.3 (9.5)	56.8 (7.7)	56.8 (7.9)
BMI, kg/m^2 (mean [SD])	24.8 (2.8)	24.7 (2.9)	22.9 (3.0)	22.8 (3.0)
Body fat percentage (mean [SD])	22.0 (5.4)	22.0 (5.3)	28.3 (5.8)	28.4 (5.6)
Waist circumference, cm (mean [SD])	88.5 (7.5)	88.4 (7.9)	80.5 (7.9)	80.4 (7.9)
Waist hip ratio (mean [SD])	0.93 (0.1)	0.93 (0.1)	0.8 (0.1)	0.8 (0.1)
Diet, %				
Non-vegetarian	4232 (96.0)	1064 (96.6)	3205 (93.8)	811 (94.9)
Vegetarian	176 (4.0)	38 (3.4)	212 (6.2)	44 (5.1)
Weekly alcohol consumption, %				
<3 times/week	3876 (87.9)	950 (86.2)	3329 (97.4)	832 (97.3)
≥3 times/week	532 (12.1)	152 (13.8)	88 (2.6)	23 (2.7)
Smoked 20 pack-years, %				
<20 pack-years	3013 (68.4)	764 (69.3)	3369 (98.6)	829 (97.0) *
≥20 pack-years	1395 (31.6)	338 (30.7)	48 (1.4)	26 (3.0)
At least 150 min of weekly exercise, %				
No	2399 (54.4)	578 (52.5)	2111 (61.8)	523 (61.2)
Yes	2009 (45.6)	524 (47.5)	1306 (38.2)	332 (38.8)
Hypertension, %				
No	2865 (65.0)	674 (61.2) *	3327 (97.4)	657 (76.8)
Yes	1543 (35.0)	428 (38.8)	734 (2.6)	198 (23.2)
Diabetes, %				
No	3691 (83.7)	929 (84.3)	3049 (89.2)	768 (89.8)
Yes	717 (16.3)	173 (15.7)	368 (10.8)	87 (10.2)
Hyperlipidemia, %				
No	3012 (68.3)	762 (69.1)	2459 (72.0)	634 (74.2)
Yes	1396 (11.7)	340 (30.9)	958 (28.0)	221 (25.8)
ASCVD, %				
No	3973 (90.1)	986 (89.5)	3327 (97.4)	821 (96.0)
Yes	435 (9.9)	116 (10.5)	90 (2.6)	34 (4.0)
Autoimmune disease, %				
No	4338 (98.4)	1084 (98.4)	3307 (96.8)	831 (97.2)
Yes	70 (1.6)	18 (1.6)	110 (3.2)	24 (2.8)
Thyroid disease, %				
No	4222 (95.8)	1056 (95.8)	3275 (95.8)	809 (94.6)
Yes	186 (4.2)	46 (4.2)	142 (4.2)	46 (5.4)
Chronic liver disease, %				
No	3716 (84.3)	933 (84.7)	3002 (87.9)	769 (89.9)
Yes	692 (15.7)	169 (15.3)	415 (12.1)	86 (10.1)
Hormone therapy, %				
No	X	X	3333 (97.5)	834 (97.5)
Yes	X	X	84 (2.5)	21 (2.5)
Albumin, g/dL	4.4 (0.3)	4.4 (0.2)	4.4 (0.3)	4.3 (0.3)
GOT, U/L	23.4 (12.6)	23.3 (13.5)	23.2 (13.6)	23.1 (10.8)
GPT, U/L	27.5 (24.0)	27.2 (19.7)	27.5 (25.4)	26.7 (18.8)
ALK-P, U/L	62.3 (13.0)	61.8 (12.8)	61.7 (11.7)	61.9 (12.2)
GGT, U/L	29.6 (26.8)	30.9 (44.5)	28.9 (23.6)	29.6 (22.4)
Total bilirubin, mg/dL	0.8 (0.4)	0.8 (0.4)	0.8 (0.4)	0.8 (0.4)
BUN, mg/dL	12.8 (3.7)	12.9 (3.6)	12.8 (4.0)	12.8 (4.1)
Creatinine, mg/dL	0.9 (0.2)	0.9 (0.2)	0.9 (0.3)	0.9 (0.2)
eGFR, mL/min/1.73 m^2	85.9 (16.2)	85.9 (16.5)	85.6 (16.3)	85.5 (16.0)
Uric acid, mg/dL	5.7 (1.4)	5.7 (1.5)	5.7 (1.4)	5.7 (1.5)
Fasting sugar, mg/dL	94.37 (20.0)	94.7 (21.6)	94.5 (19.8)	95.8 (20.8)
HbA1c, %	5.8 (0.7)	5.8 (0.8)	5.8 (0.7)	5.9 (0.7)
Total cholesterol, mg/dL	195.7 (35.9)	196.5 (40.1)	195.0 (37.5)	195.1 (36.2)
HDL, mg/dL	53.4 (13.8)	53.5 (14.0)	53.1 (13.2)	53.6 (14.3)
LDL, mg/dL	118.4 (31.8)	118.6 (34.1)	117.2 (31.1)	116.2 (32.0)
TG, mg/dL	122.9 (99.7)	123.7 (90.0)	123.1 (95.6)	122.3 (90.1)
Osteoporosis, %				
No	3758 (85.3)	951 (86.3)	2306 (67.5)	578 (67.6)
Yes	650 (14.7)	151 (13.7)	1111 (32.5)	277 (32.4)

Values are presented as numbers (%) or mean (standard deviation [SD]).
*Significant difference in Chi-square test and t-test at $P < 0.05$, **$P < 0.01$, and ***$P < 0.001$.
HbA1c: glycohemoglobin; HDL: high-density lipoprotein cholesterol; LDL: low-density lipoprotein cholesterol; TG: triglycerides; GOT: glutamate oxaloacetate transaminase; GPT: glutamate pyruvate transaminase; ALK-P: alkaline phosphatase; BUN: blood urea nitrogen; eGFR: estimated glomerular filtration rate; GGT: gamma-glutamyl transferase; ASCVD: atherosclerotic cardiovascular disease, BMI: body mass index.

of the models ranged from 0.59 to 0.77, with the ANN model exhibiting the highest AUC value of 0.77 (95 % CI: 0.729–0.801). The KNN and SVM models followed, with AUC values of 0.64 (95 % CI: 0.603–0.685) and 0.63 (95 % CI: 0.585–0.667), respectively (Fig. 2). Similar to men, in women, the OSTA model had the lowest AUC value of 0.55 (95 % CI: 0.536–0.572), with the highest specificity of 0.97.

The SVM, XGB, and ANN models in men and the DT, RF, SVM, KNN, XGB, and ANN models in women performed significantly better than the OSTA model did. The results of this study suggest that the ANN model is the most promising candidate for predicting osteoporosis among women in the Taiwanese population, as it demonstrated the highest AUC.

We gathered an additional 1133 datasets for external validation to validate the predictive performance of the models. The ANN model demonstrated similar AUC values of 0.70 (95 % CI: 0.643–0.760) and 0.71 (95 % CI: 0.665–0.762) in men and women, respectively (Table 4); the ROC curve is demonstrated in Supplementary Fig. 1.

4. Discussion

We explored a novel approach, using various ML models, to predict the risk of osteoporosis with four simple clinical features, sex, age, height, and weight, in postmenopausal women and men aged ≥50 years in southern Taiwan. The results of this study suggest that the ANN model is the most promising candidate for predicting osteoporosis, as it demonstrated the highest AUC and performed better, compared with the OSTA model.

Our study has several strengths. First, to our knowledge, this is the simplest ML model for predicting osteoporosis. According to a review by Smets et al. (2021) most existing ML models for osteoporosis prediction depend on high-dimensional and complex inputs, including numerous clinical features and medical images. Although these models achieve high accuracy, they are often less practical for widespread use. Our approach significantly reduces complexity and computational demands,

Table 3
Performance of all machine learning models, stratified by sex.

ML model (men)	AUC (95 % CI)	Sensitivity	Specificity	Accuracy	*p*-Value (vs. OSTA)	ML model (women)	AUC (95 % CI)	Sensitivity	Specificity	Accuracy	p-Value (vs. OSTA)
OSTA	0.51 (0.493–0.535)	0.03	0.99	0.85	X	OSTA	0.55 (0.536–0.572)	0.14	0.97	0.71	X
DT	0.52 (0.473–0.567)	0.20	0.84	0.74	0.22	DT	0.59 (0.544–0.627)	0.47	0.71	0.63	0.04
RF	0.52 (0.472–0.565)	0.08	0.96	0.82	0.08	RF	0.62 (0.581–0.662)	0.41	0.83	0.70	<0.001
SVM	0.55 (0.506–0.601)	0.54	0.57	0.54	0.01	SVM	0.63 (0.585–0.667)	0.69	0.56	0.60	<0.001
KNN	0.51 (0.466–0.560)	0.06	0.96	0.82	0.19	KNN	0.64 (0.603–0.685)	0.45	0.84	0.71	<0.001
XGB	0.53 (0.482–0.575)	0.09	0.97	0.83	0.01	XGB	0.62 (0.576–0.658)	0.40	0.83	0.69	<0.001
ANN	0.67 (0.622–0.715)	0.22	0.92	0.81	0.03	ANN	0.77 (0.729–0.801)	0.55	0.82	0.73	<0.001

OSTA: Osteoporosis Self-Assessment Tool for Asians; DT: decision trees; RT: random forests; SVM: support vector machines; KNN: k-nearest neighbors; XGB: extreme gradient boosting; ANN: artificial neural networks; AUC: area under the receiver operating characteristic; CI: confidence interval; ML: machine learning.

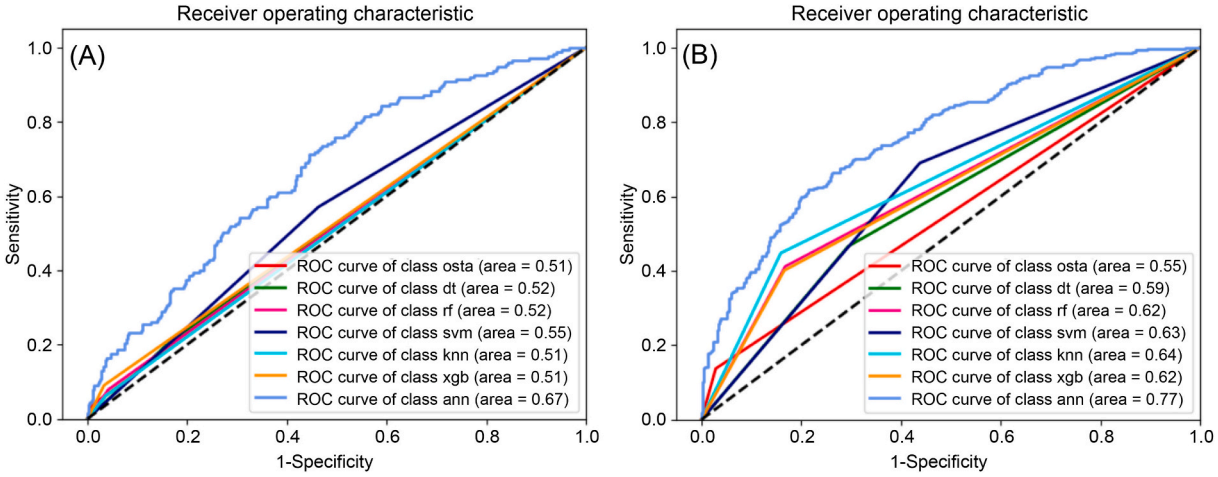

Fig. 2. ROC curves for ML models to predict osteoporosis.
Models in men (A) and women (B) compared to OSTA. ROC: receiver operating characteristics; ML: machine learning; OSTA: Osteoporosis Self-Assessment Tool for Asia.

Table 4
Performance of artificial neural network model using the validation dataset (n = 1133).

Sex	AUC (95 % CI)	Sensitivity	Specificity	Accuracy
Men	0.70 (0.643–0.760)	0.25	0.90	0.79
Women	0.71 (0.665–0.762)	0.53	0.77	0.69

AUC: area under the receiver operating characteristic; CI: confidence interval.

making it accessible to non-specialists, including the general public. This simplification enhances feasibility for population screening and can allow for the identification of high-risk individuals for early intervention treatments. Second, we utilized DXA scans, the gold standard for measuring BMD, to assess the BMD of the lumbar spine and bilateral hips (Kanis and Glüer, 2000). This approach, combined with our study conducted across a large-scale population in Taiwan, greatly enhances the precision and reliability of our findings. Third, we employed six ML models (DT, RF, SVM, KNN, XGB, and ANN) currently in use and incorporated various clinical features. Additionally, we included an extra validation dataset for sensitivity analysis, ensuring the robustness of our results across different populations (Smets et al., 2021).

Our findings are consistent with those of previous studies, highlighting age, height, weight, and BMI as the primary factors associated with osteoporosis (Yoo et al., 2013; Shim et al., 2020; Ou Yang et al., 2021; Wu and Park, 2023). Shim et al. (2020) created an ML model for osteoporosis prediction in 1792 postmenopausal women, and observed optimal performance in the ANN model. Our study differs from the approach used by Shim et al. of including broader variables and not including comparisons with conventional models, such as the OSTA model. They selected more variables using logistic regression, such as the duration of menopause and history of fracture. They also developed another model that included all variables directly into the analysis, and the results indicated comparable predictive capabilities between the two models. Another study by Yoo et al. (2013), which focused on 1674 postmenopausal women, developed three ML models (SVM, RF, and ANN) to predict osteoporosis. Notably, all three models incorporated age, weight, and BMI as variables. Due to the use of ML methods, various epidemiological factors of osteoporosis could be considered, leading to the ML models demonstrating superior performance, compared with the OSTA model. Among the ML models, SVM emerged as the most effective based on the AUC. Notably, both studies primarily concentrated on osteoporosis prediction in postmenopausal women. However, our study developed a model that predicts osteoporosis in both men and women. Regarding studies including both sexes, Wu and Park (2023) conducted a study involving 4037 men and 4385 women aged ≥40 years, developing an ML model using the XGB method with 15 input features. However, they measured BMD using QUS at the radius or tibia and the models did not differentiate between men and women. Considering the

distinct etiology of osteoporosis in men, where secondary osteoporosis is more common (Vilaca et al., 2022), and in women, where it is associated with postmenopausal estrogen deficiency, our approach involved separate model training, acknowledging the varying causes and prevalence of osteoporosis in men and women. Another study conducted in northern Taiwan by Ou Yang et al. (2021) encompassed 3053 men and 2929 women aged ≥50 years, utilizing DXA for BMD measurement. They developed four ML models (RF, KNN, SVM, ANN) with 16 input features for men and 19 for women. They developed separate models for men and women; however, the observed prevalence of osteoporosis in their study was notably lower, at 3.8 % and 10.4 % in men and women, compared with local population rates of 13.3 % and 36.0 % in men and women (Chen et al., 2018), and our observed rates of 14.5 % and 32.5 % in men and women, respectively. This discrepancy shows potential differences in the study cohort compared with the broader population and raises concerns that the findings may not accurately represent the authentic demographic profile. Suh et al. (2023) employed a deep learning approach for osteoporosis risk screening in 8274 Americans (51.8 % male) and 8680 Koreans (44.9 % male). The deep learning approach had superior performance, compared with traditional ML models, aligning with our findings that identified age, sex, and BMI as crucial variables. However, the complexity of these deep learning models, with several input features, remains a challenge for general population use and clinical implementation.

Based on cutting-edge ML techniques, our model continuously refines itself with incoming data, enabling sophisticated analysis and interpretation. This dynamic updating capability allows it to capture intricate patterns that traditional models, such as the OSTA model, might overlook. Furthermore, incorporating an additional validation dataset for sensitivity analysis further strengthened the reliability of our findings across diverse populations. This ensures that our model's performance is robust and applicable to several individuals, contributing to the broader validity and generalizability of our results.

This study has some limitations. First, it adopted a cross-sectional investigation, which may not fully represent the condition of the actual general population. However, the study's substantial sample size mitigated this concern, and its prevalence of osteoporosis aligned with previous epidemiological findings (Chen et al., 2018). Second, our study lacked validation across diverse populations. Therefore, further research incorporating heterogeneous populations is necessary to verify our findings. Incorporating diverse data from various populations may allow for refinement of the model, potentially enhancing its generalizability and predictive power across different settings. Third, specific factors, such as disease duration, fracture history, medication use (e.g., hormone therapy and steroids), and the exact daily amount or specific units of alcohol consumed, were not included in the ML models due to the lack of precise data. Future research could incorporate the aforementioned variables to predict the risk of osteoporosis and fractures.

In conclusion, this study's primary finding is facilitating the identification of postmenopausal women and men aged ≥50 years at high risk of osteoporosis in southern Taiwan. The ANN model surpasses other ML models in predicting osteoporosis risk using a limited set of variables, outperforming even the OSTA model, which also relies on simple variables. The general public can readily use this model to initiate preventive and therapeutic measures for osteoporosis in Taiwan. However, further studies should include extended populations and extensively explore the severity of osteoporosis.

Abbreviations

ASCVD	atherosclerotic cardiovascular disease
BMD	bone mineral density
DM	diabetes mellitus
DXA	dual-energy X-ray absorptiometry
DTs	decision trees
HTN	hypertension
KNNs	k-nearest neighbors
XGB	extreme gradient boosting
ANNs	artificial neural networks
ML	machine learning
OSTA	Osteoporosis Self-Assessment Tool for Asians
ROC	receiver operating characteristic
QUS	quantitative ultrasound
RFs	random forests
SVMs	support vector machines
WC	waist circumference
WHR	waist-hip ratio

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bonr.2025.101826.

CRediT authorship contribution statement

Wei-Chin Huang: Writing – original draft, Validation, Investigation, Formal analysis, Data curation. **I-Shu Chen:** Software, Resources, Project administration. **Hsien-Chung Yu:** Software, Resources, Project administration. **Chi-Shen Chen:** Investigation, Data curation. **Fu-Zong Wu:** Validation, Formal analysis. **Chiao-Lin Hsu:** Writing – review & editing, Supervision, Methodology, Investigation, Funding acquisition, Conceptualization. **Pin-Chieh Wu:** Writing – review & editing, Supervision, Methodology, Investigation, Conceptualization.

Funding

This work was supported by the Kaohsiung Veterans General Hospital [grant number KSVGH 112-132].

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

The authors would like to thank Kaohsiung Veterans General Hospital for providing the anonymous data.

Appendix A

$$\text{estimated glomerular filtration rate } \left(\text{mL}/min/1.73\ \text{m}^2\right) = 175 \times (\text{serum creatinine}) - 1.154 \times (\text{age, years}) - 0.203 \times (0.742 \text{ if female}) \quad (1)$$

Data availability

This study includes ongoing research data, which is temporarily unavailable for sharing.

References

Chai, H., Ge, J., Li, L., Li, J., Ye, Y., 2021. Hypertension is associated with osteoporosis: a case-control study in Chinese postmenopausal women. BMC Musculoskelet. Disord. 22, 253. https://doi.org/10.1186/s12891-021-04124-9.

Chen, F.P., Huang, T.S., Fu, T.S., Sun, C.C., Chao, A.S., Tsai, T.L., 2018. Secular trends in incidence of osteoporosis in Taiwan: a nationwide population-based study. Biom. J. 41, 314–320. https://doi.org/10.1016/j.bj.2018.08.001.

Consensus Development Conference, 1993. Consensus development conference: diagnosis, prophylaxis, and treatment of osteoporosis. Am. J. Med. 94, 646–650. https://doi.org/10.1016/0002-9343(93)90218-e.

Delitala, A.P., Scuteri, A., Doria, C., 2020. Thyroid hormone diseases and osteoporosis. J. Clin. Med. 9, 1034. https://doi.org/10.3390/jcm9041034.

Drinkwater, B.L., 1994. McCloy Research Lecture: does physical activity play a role in preventing osteoporosis? Res. Q. Exerc. Sport 65, 197–206. https://doi.org/10.1080/02701367.1994.10607620.

Ensrud, K.E., Crandall, C.J., 2017. Osteoporosis. Ann. Intern. Med. 167, ITC17–ITC32. https://doi.org/10.7326/AITC201708010.

Godos, J., Giampieri, F., Chisari, E., Micek, A., Paladino, N., Forbes-Hernández, T.Y., Quiles, J.L., Battino, M., La Vignera, S., Musumeci, G., Grosso, G., 2022. Alcohol consumption, bone mineral density, and risk of osteoporotic fractures: a dose–response meta-analysis. Int. J. Environ. Res. Public Health 19, 1515. https://doi.org/10.3390/ijerph19031515.

Goldring, S.R., Gravallese, E.M., 2000. Mechanisms of bone loss in inflammatory arthritis: diagnosis and therapeutic implications. Arthritis Res. 2, 33–37. https://doi.org/10.1186/ar67.

Kanis, J.A., Glüer, C.C., 2000. An update on the diagnosis and assessment of osteoporosis with densitometry. Committee of Scientific Advisors, International Osteoporosis Foundation. Osteoporos. Int. 11, 192–202. https://doi.org/10.1007/s001980050281.

Koh, L.K., Sedrine, W.B., Torralba, T.P., Kung, A., Fujiwara, S., Chan, S.P., Huang, Q.R., Rajatanavin, R., Tsai, K.S., Park, H.M., Reginster, J.Y., Osteoporosis Self-Assessment Tool for Asians (OSTA) Research Group, 2001. A simple tool to identify Asian women at increased risk of osteoporosis. Osteoporos. Int. 12, 699–705. https://doi.org/10.1007/s001980170070.

Krall, E.A., Dawson-Hughes, B., 1999. Smoking increases bone loss and decreases intestinal calcium absorption. J. Bone Miner. Res. 14, 215–220. https://doi.org/10.1359/jbmr.1999.14.2.215.

Ou Yang, W.Y., Lai, C.C., Tsou, M.T., Hwang, L.C., 2021. Development of machine learning models for prediction of osteoporosis from clinical health examination data. Int. J. Environ. Res. Public Health 18, 7635. https://doi.org/10.3390/ijerph18147635.

Pirih, F., Lu, J., Ye, F., Bezouglaia, O., Atti, E., Ascenzi, M.G., Tetradis, S., Demer, L., Aghaloo, T., Tintut, Y., 2012. Adverse effects of hyperlipidemia on bone regeneration and strength. J. Bone Miner. Res. 27, 309–318. https://doi.org/10.1002/jbmr.541.

Shim, J.G., Kim, D.W., Ryu, K.H., Cho, E.A., Ahn, J.H., Kim, J.I., Lee, S.H., 2020. Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women. Arch. Osteoporos. 15, 169. https://doi.org/10.1007/s11657-020-00802-8.

Smets, J., Shevroja, E., Hügle, T., Leslie, W.D., Hans, D., 2021. Machine learning solutions for osteoporosis-a review. J. Bone Miner. Res. 36, 833–851. https://doi.org/10.1002/jbmr.4292.

Suh, B., Yu, H., Kim, H., Lee, S., Kong, S., Kim, J.W., Choi, J., 2023. Interpretable deep-learning approaches for osteoporosis risk screening and individualized feature analysis using large population-based data: model development and performance evaluation. J. Med. Internet Res. 25, e40179. https://doi.org/10.2196/40179.

Tao, J., Yan, Z., Huang, W., Feng, T., 2023. Seropositive for hepatitis B and C viruses is associated with the risk of decreased bone mineral density in adults: an analysis of studies from the NHANES database. Front. Med (Lausanne) 10, 1120083. https://doi.org/10.3389/fmed.2023.1120083.

Vilaca, T., Eastell, R., Schini, M., 2022. Osteoporosis in men. Lancet. Diabetes and Endocrinology 10, 273–283. https://doi.org/10.1016/S2213-8587(22)00012-2.

Wang, C.Y., Fu, S.H., Yang, R.S., Shen, L.J., Wu, F.L.L., Hsiao, F.Y., 2017. Age- and gender-specific epidemiology, treatment patterns, and economic burden of osteoporosis and associated fracture in Taiwan between 2009 and 2013. Arch. Osteoporos. 12, 92. https://doi.org/10.1007/s11657-017-0385-5.

Wu, X., Park, S., 2023. A prediction model for osteoporosis risk using a machine-learning approach and its validation in a large cohort. J. Korean Med. Sci. 38, e162. https://doi.org/10.3346/jkms.2023.38.e162.

Yang, N.P., Deng, C.Y., Chou, Y.J., Chen, P.Q., Lin, C.H., Chou, P., Chang, H.J., 2006. Estimated prevalence of osteoporosis from a Nationwide Health Insurance database in Taiwan. Health Policy 75, 329–337. https://doi.org/10.1016/j.healthpol.2005.04.009.

Yang, J.J., Yen, H.K., Li, C.Y., Chen, H.J., Liou, H.H., Fu, S.H., Wang, C.Y., 2023. Ten existing osteoporosis prediction tools for the successful application of National Health Insurance-reimbursed anti-osteoporosis medications in long-term care residents in Taiwan. J. Formos. Med. Assoc. 122, 139–147. https://doi.org/10.1016/j.jfma.2022.08.009.

Yaturu, S., 2009. Diabetes and skeletal health. J. Diabetes 1, 246–254. https://doi.org/10.1111/j.1753-0407.2009.00049.x.

Yoo, T.K., Kim, S.K., Kim, D.W., Choi, J.Y., Lee, W.H., Oh, E., Park, E.C., 2013. Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. Yonsei Med. J. 54, 1321–1330. https://doi.org/10.3349/ymj.2013.54.6.1321.

Glossary

Atherosclerotic cardiovascular disease: myocardial infarction, acute coronary syndrome, stable or unstable angina, stroke, transient ischemic attack, carotid disease, peripheral artery disease, and abdominal aortic aneurysm.