



## OPEN Machine learning models to predict osteoporosis in patients with chronic kidney disease stage 3–5 and end-stage kidney disease

Chia-Tien Hsu<sup>1,2,3,4</sup>, Chin-Yin Huang<sup>1✉</sup>, Cheng-Hsu Chen<sup>2,5</sup>, Ya-Lian Deng<sup>4,6</sup>, Shih-Yi Lin<sup>3,4,7</sup> & Ming-Ju Wu<sup>2,5</sup>

Chronic kidney disease-mineral bone disorder is a common complication in patients with chronic kidney disease (CKD) and end-stage kidney disease (ESKD), and it increases the risk of osteoporosis and fractures. This study aimed to develop predictive machine-learning (ML) models to identify osteoporosis risk in patients with CKD stages 3–5 and ESKD. We retrospectively analyzed a de-identified osteoporosis database from a Taiwanese hospital, including 6614 patients with CKD stages 3–5 and ESKD who underwent bone mineral density (BMD) scans between January 2011 and June 2022. Nine ML algorithms were applied to predict osteoporosis: logistic regression, XGBoost, LightGBM, CatBoost, SVM, decision tree, random forest, k-nearest neighbors, and an artificial neural network (ANN). The ANN model achieved the highest predictive performance, with an area under the curve (AUC) of 0.940 on the validation and 0.930 on the test datasets. The receiver operating characteristic curve, confusion matrix, and predictive probability histogram revealed that the ANN model performed well in terms of discrimination. Calibration and decision curve analyses further demonstrated the reliability and applicability of the ANN model. The ANN model demonstrated the potential for clinical implementation in screening high-risk patients for osteoporosis.

**Keywords** Machine learning, Osteoporosis, Chronic kidney disease, End-stage kidney disease, Bone mineral density

Chronic kidney disease-mineral bone disorder (CKD-MBD) is one of the most common complications in patients with chronic kidney disease (CKD) and end-stage kidney disease (ESKD). The impact of CKD-MBD is increasingly being recognized with the global rise in CKD<sup>1</sup>. Skeletal abnormalities associated with CKD-MBD lead to a significantly higher risk of osteoporosis and fractures. Studies show that 18–32% of patients with CKD have osteoporosis<sup>2</sup>, with a fracture risk over 2.5 times higher than in those without CKD. For patients with ESKD, the risk was more than four times higher. Fragility fractures in patients with advanced CKD and ESKD lead to increased morbidity, mortality, and healthcare costs<sup>2–6</sup>.

Bone biopsy is considered the gold standard for diagnosing CKD-MBD; however, its invasiveness and cost limit its use. The 2017 Kidney Disease Improving Global Outcomes (KDIGO) guidelines recommend that bone densitometry is essential for assessing fracture risk in patients with CKD, as disruptions in bone and mineral metabolism caused by CKD increase fracture risk<sup>7</sup>. Osteoporosis is diagnosed by a T-score of  $\leq -2.5$  at the lumbar spine, femur neck, or total hip, using bone mineral density (BMD) testing via dual-energy X-ray absorptiometry (DXA)<sup>8</sup>. DXA is the most widely used method to measure BMD and diagnose osteoporosis. However, in clinical practice, many patients with CKD do not undergo BMD testing. In Taiwan, DXA scan access is limited by healthcare coverage restrictions<sup>9</sup>, and machines are not available in all clinical settings because of their cost

<sup>1</sup>Department of Industrial Engineering and Enterprise Information, Tunghai University, Taichung 407224, Taiwan.

<sup>2</sup>Division of Nephrology, Department of Internal Medicine, Taichung Veterans General Hospital, Taichung 407219, Taiwan. <sup>3</sup>School of Medicine, National Yang Ming Chiao Tung University, Taipei 112304, Taiwan. <sup>4</sup>Center for Osteoporosis Prevention and Treatment, Taichung Veterans General Hospital, Taichung 407219, Taiwan.

<sup>5</sup>Department of Post-Baccalaureate Medicine, College of Medicine, National Chung Hsing University, Taichung 40227, Taiwan. <sup>6</sup>Department of Nursing, Taichung Veterans General Hospital, Taichung 407219, Taiwan.

<sup>7</sup>Center for Geriatrics and Gerontology, Taichung Veterans General Hospital, Taichung 407219, Taiwan. ✉email: huangcy@thu.edu.tw

and space requirements. Therefore, tools are required to screen patients with CKD who should undergo DXA scanning.

Artificial intelligence has been widely employed to assist in disease diagnosis and outcome prediction. Machine learning (ML) models are mathematical algorithms used to identify patterns in data<sup>10,11</sup>. Several studies have explored ML applications for predicting osteoporosis, focusing on diverse populations, including the general population<sup>12–20</sup>, postmenopausal women<sup>21–24</sup>, and patients with breast cancer<sup>25</sup>, rheumatoid arthritis<sup>26</sup>, and type 2 diabetes mellitus<sup>27,28</sup>. However, no studies have employed multiple ML models to predict osteoporosis in patients with CKD stage 3–5 and ESKD. Therefore, our study aimed to apply various ML models to predict osteoporosis in this population and develop a one-year predictive ML model for osteoporosis risk.

## Methods

### Study participants

Our hospital has established a de-identified osteoporosis database that includes all participants who have undergone BMD scans. A total of 35,670 BMD scans were recorded in the database between January 1, 2011, and June 30, 2022. Demographic information, comorbidities, laboratory and BMD scan data were retrospectively extracted from the database. The study population selected from the database comprised participants with CKD stages 3–5 and ESKD who had undergone at least one BMD scan. We primarily enrolled participants with CKD and ESKD with at least two tests, separated by at least 3 months, showing estimated glomerular filtration rate (eGFR) < 60 mL/min/1.73 m<sup>2</sup>. Additionally, one of the two eGFR qualifying records had to be within one year prior to the BMD scan. We excluded participants who did not have a renal function test within one year prior to the BMD scan. A total of 6614 eligible participants were included in our cohort for further analysis. This study was approved by the Institutional Review Board of the Taichung Veterans General Hospital (No. SE24210B). All methods were carried out following the relevant institutional guidelines and regulations. The Institutional Review Board of the Taichung Veterans General Hospital approved a waiver of informed consent, as all protected health information was de-identified, and the study was retrospective in nature.

### Data extraction

All extracted patient information was de-identified. The demographic features used for the machine-learning models included age, sex, height, weight, and body mass index (BMI). Laboratory data included eGFR, serum creatinine (Cre), serum albumin, serum corrected calcium (Ca), serum phosphorus, serum alkaline phosphatase (ALP), serum intact parathyroid hormone (i-PTH), white blood cell count (WBC), hemoglobin (HGB), fasting glucose, and glycated hemoglobin. Comorbidities were identified using ICD-9 or ICD-10 codes and included osteoarthritis, rheumatoid arthritis, diabetes mellitus, hypertension, cerebrovascular accident, cataracts, and a history of previous fractures.

Additionally, we extracted data on the comorbidities related to secondary osteoporosis, including type 1 diabetes mellitus, osteogenic imperfecta in adults, hyperthyroidism, hypogonadism, premature ovarian insufficiency, chronic malnutrition, and chronic liver disease. The comorbidities related to secondary osteoporosis were extracted from the Fracture Risk Assessment Tool (FRAX®)<sup>29</sup> evaluations of these patients, which consider disorders associated with secondary osteoporosis. The assessment was conducted by our nurse case managers. Patients were classified as having chronic malnutrition if they had a diagnosis of ICD-10 codes E40–E46 or ICD-9 codes 260–263, or if our nurse case managers determined that they met the criteria for protein-energy wasting (PEW). The diagnosis of PEW was based on the definition established by the International Society of Renal Nutrition and Metabolism<sup>30</sup>. History of steroid use was also extracted if the patient had been taking oral glucocorticoids for more than 3 months at a daily dose of at least 5 mg of prednisolone. EGFR was calculated using the 2021 Chronic Kidney Disease Epidemiology Collaboration creatinine equation<sup>31,32</sup>. Corrected calcium levels in the serum were estimated using the following formula: corrected calcium (mg/dL) = serum calcium (mg/dL) + 0.8 × (4 – serum albumin [g/dL])<sup>33</sup>.

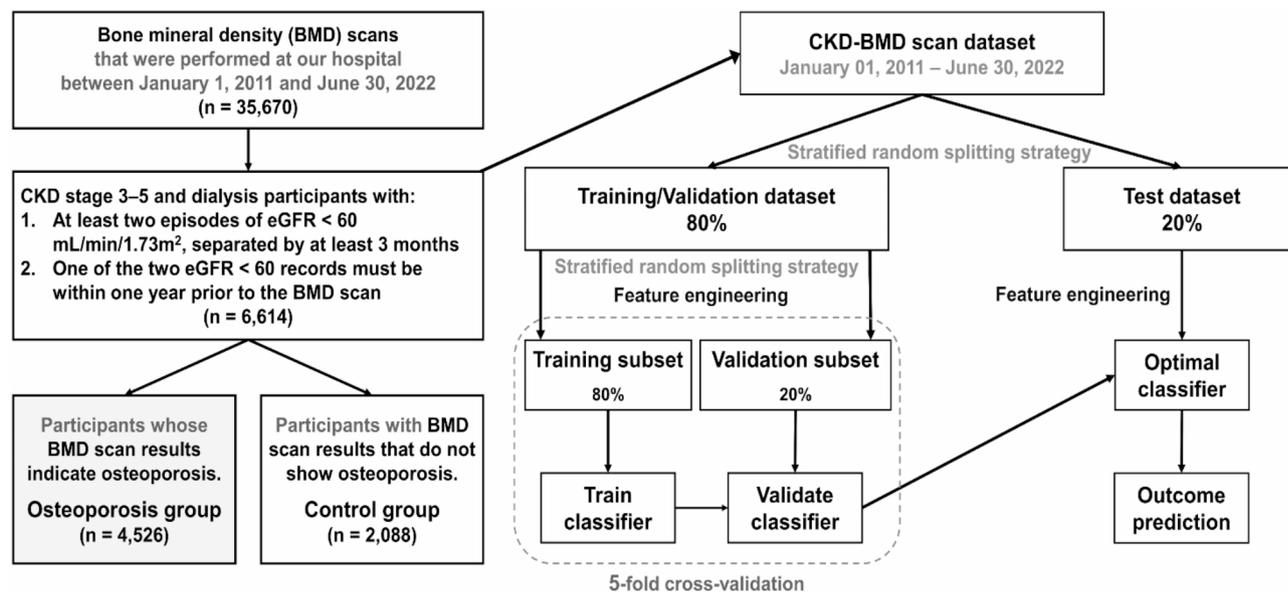
### Study design and label definition

Figure 1 illustrates the study design, label definitions, and architectures of the prediction models. According to the World Health Organization's operational definition, the diagnosis of osteoporosis is based on a clinical history of fragility fractures or a T-score of ≤ –2.5 at the lumbar spine, femoral neck, or total hip, as determined by BMD testing using DXA<sup>8</sup>. Participants were labeled as the “osteoporosis” group if their BMD scan results indicated osteoporosis. Conversely, participants were labeled as the “control” group if their BMD results did not show osteoporosis.

We used a stratified random splitting strategy to divide the study cohort into 80% training and validation data and 20% test data. This strategy ensured that the proportion of osteoporosis cases was the same in both the training and test sets<sup>34</sup>. Next, we applied the same stratified random splitting strategy to further divide the training and validation data into 80% training and 20% validation subsets. To identify the optimal classifier, we performed five-fold cross-validation of the training and validation subsets. The optimal classifier, independent of the training/validation data, was used to predict the outcomes for each participant in the test dataset. This provided an unbiased final performance metric for the model<sup>35</sup>.

### Data preprocessing and feature selection

Table 1 presents the missing data of the study cohort. We conducted Little's missing completely at random (MCAR) test, which resulted in a *P*-value > 0.05, indicating that the missing data pattern did not meet the MCAR assumption<sup>36</sup>. In other words, the data are not missing completely at random and may instead be missing at random or missing not at random. For example, the probability of missing BMI data is related to the height and weight features<sup>37</sup>. We conducted a pilot study using three different methods to impute missing values:



**Fig. 1.** Study design, label definitions, and the architecture of our prediction models.

mean imputation, multiple imputation, and K-nearest neighbors (KNN) imputation<sup>38</sup>. The results showed that multiple imputations yielded fair performance; therefore, we selected it as the final method for handling missing values. Continuous variables were standardized using standard scaling prior to training. Categorical variables were processed using one-hot encoding prior to training.

Next, we applied recursive feature elimination with a five-fold cross-validation (RFECV) to determine the optimal number of features<sup>39</sup>. The RFECV results indicate that 16 features were optimal. We reviewed the literature on ML applications for predicting osteoporosis<sup>12–28</sup> and applied three methods to select features and gradually reduce the number of features. Table 2 outlines the three feature-selection methods and the features chosen in our study. In the first method, features are selected based on their importance<sup>17,18,23,24</sup>. In the second method, we used recursive feature elimination (RFE) with a random forest classifier to identify important features. RFE is a backward feature selection technique that optimizes the model performance by iteratively removing the least important features based on their predictive values<sup>20,24</sup>. In the third method, key features were selected from those identified in methods 1 and 2 through group discussions. Additionally, during the feature-selection process in method 3, we considered the accessibility and practicality of the selected features for clinical use<sup>12,13</sup>.

### Machine learning models and performance evaluation

We applied nine ML algorithms to predict osteoporosis in patients with CKD stages 3–5 and ESKD: logistic regression (LR), extreme gradient boosting (XGBoost), light gradient boosting machine (LightGBM), category boosting (CatBoost), support vector machine (SVM), decision tree (DT), random forest (RF), KNN, and artificial neural networks (ANN). We evaluated the performance of each model using accuracy, precision, recall, F1 score, and area under the curve (AUC). In this study, the AUC refers to the area under the receiver operating characteristic (ROC) curve. The 95% confidence interval (CI) for each endpoint was computed using the bootstrap method, where 1000 replacement resampling procedures were performed to evaluate the model performance. We also used ROC curve, calibration curve, and decision curve analysis to determine the discrimination, reliability, and applicability of the predictive ML models. Additionally, we calculated the Shapley additive explanations (SHAP) values to assess feature importance and explore the relationship between outcomes and features<sup>40</sup>. All analyses, including data preprocessing, feature selection, ML modeling, and performance evaluation, were conducted using Python (version 3.12.4)<sup>41</sup>.

In this study, we employed different hyperparameter tuning approaches based on the characteristics of each machine learning model. For the ANN model, we performed grid search with threefold cross-validation to determine the optimal hyperparameters, exploring batch sizes (32, 64, 128, 256), epochs (50, 100, 200, 300), dropout rates (0.05, 0.1, 0.2, 0.3), and optimizers (Adam, RMSprop). Early stopping was implemented to monitor validation loss and halt training if no improvement was observed for 10 consecutive epochs. For other machine learning models, we used randomized search with fivefold cross-validation to efficiently explore the hyperparameter space in models such as XGBoost, LightGBM, and random forest, while for models like logistic regression, decision tree, KNN, and SVM, we relied on default Scikit-learn settings to ensure standard implementation and baseline comparisons. The final hyperparameter settings for each model are summarized in Supplemental Table S1.

Normality was assessed using the Kolmogorov–Smirnov test to compare the demographic and clinical characteristics of the participants. Tests for statistical significance were conducted using the Mann–Whitney U test for continuous variables and the chi-square test or Fisher’s exact test for categorical variables. The level

Variable	Overall (n = 6614)		Osteoporosis Group (n = 4526)		Control Group (n = 2088)		P-Value
	Number	Percentage	Number	Percentage	Number	Percentage	
Age	0	0%	0	0%	0	0%	–
Female sex	0	0%	0	0%	0	0%	–
Height	2288	34.59%	1852	40.92%	436	20.88%	<0.001
Weight	2048	30.96%	1676	37.03%	372	17.82%	<0.001
BMI	2303	34.82%	1863	41.16%	440	21.07%	<0.001
eGFR	0	0%	0	0%	0	0%	–
Serum creatinine	0	0%	0	0%	0	0%	–
Serum albumin	669	10.11%	457	10.10%	212	10.15%	0.944
ALP	1072	16.21%	729	16.11%	343	16.43%	0.743
Corrected calcium	472	7.14%	307	6.78%	165	7.90%	0.100
Phosphorus	2092	31.63%	1442	31.86%	650	31.13%	0.553
i-PTH	3804	57.51%	2684	59.30%	1120	53.64%	<0.001
WBC count	1545	23.36%	1107	24.46%	438	20.98%	0.002
HGB	1526	23.07%	1076	23.77%	450	21.55%	0.046
HbA1c	1323	20.00%	1031	22.78%	292	13.98%	<0.001
Fasting Glucose	589	8.86%	419	9.26%	167	8.00%	0.094
Previous fracture	0	0%	0	0%	0	0%	–
OA	0	0%	0	0%	0	0%	–
RA	0	0%	0	0%	0	0%	–
DM	0	0%	0	0%	0	0%	–
HTN	0	0%	0	0%	0	0%	–
CVA	0	0%	0	0%	0	0%	–
Cataract	0	0%	0	0%	0	0%	–
T1DM	0	0%	0	0%	0	0%	0%
Hyperthyroidism	0	0%	0	0%	0	0%	–
Hypogonadism	0	0%	0	0%	0	0%	–
POI	0	0%	0	0%	0	0%	–
Malnutrition	0	0%	0	0%	0	0%	–
CLD	0	0%	0	0%	0	0%	–
Steroid usage	0	0%	0	0%	0	0%	–

**Table 1.** Missing data in our study cohort. The *P*-value represents comparison between the osteoporosis and control groups. ALP, alkaline phosphatase; BMI, body mass index; CLD, chronic liver disease CVA, cerebrovascular accident; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate calculated using the 2021 Chronic Kidney Disease Epidemiology Collaboration creatinine equation; HbA1c, glycated hemoglobin; HGB, hemoglobin; HTN, hypertension; i-PTH, intact parathyroid hormone; OA, osteoarthritis; POI, premature ovarian insufficiency; RA, rheumatoid arthritis; T1DM, type 1 diabetes mellitus; WBC, white blood cell.

of significance was set at a two-tailed value of  $P < 0.05$ . Statistical analyses were performed using MedCalc for Windows, version 22.303 (MedCalc Software; Ostend, Belgium).

## Results

### Demographic and clinical characteristics of the participants

A total of 6614 participants with stages 3–5 CKD and ESKD were enrolled to predict osteoporosis. Among them, 4526 participants were classified into the osteoporosis group, while 2088 participants were in the control group. Table 3 presents the baseline demographic and clinical characteristics of participants. The osteoporosis group, compared to the control group, was significantly older, with a higher proportion of females, and with shorter height, lower weight, lower BMI, lower eGFR, lower serum Cre levels, lower serum albumin levels, higher ALP levels, lower corrected Ca levels, lower serum phosphorus levels, lower i-PTH levels, higher WBCs, lower HGB levels, higher fasting glucose levels, and a higher proportion of participants with a history of osteoporosis, cerebrovascular accidents, cataracts, hypogonadism, malnutrition, or chronic liver disease. Although higher Cre levels are typically associated with lower eGFR, eGFR is calculated using age, sex, and serum Cre levels. Therefore, despite the lower serum Cre levels in the osteoporosis group, the eGFR was lower in this group because of their older age and higher proportion of females.

Number of selected features	Method 1: Use “Permutation Importance” to identify important features
6	Age, previous fracture, height, weight, glucose, HbA1c
7	Age, previous fracture, height, weight, glucose, HbA1c, WBC
8	Age, previous fracture, height, weight, glucose, HbA1c, WBC, ALP
9	Age, previous fracture, height, weight, glucose, HbA1c, WBC, ALP, HGB
10	Age, previous fracture, height, weight, glucose, HbA1c, WBC, ALP, HGB, i-PTH
15	Cre, eGFR, age, weight, height, BMI, albumin, Ca, phosphate, ALP, HbA1c, glucose, WBC, HGB, previous fracture
16	Cre, eGFR, age, weight, height, BMI, albumin, Ca, phosphate, ALP, i-PTH, HbA1c, glucose, WBC, HGB, previous fracture
32	All eligible features
Number of selected features	Method 2: Use “Recursive Feature Elimination” with Random Forest Classifier to select important features
6	eGFR, age, weight, height, ALP, glucose
7	eGFR, age, weight, height, ALP, glucose, WBC
8	eGFR, age, weight, height, ALP, glucose, WBC, HGB
9	eGFR, age, weight, height, albumin, ALP, glucose, WBC, HGB
10	eGFR, Cre, age, weight, height, albumin, ALP, glucose, WBC, HGB
15	eGFR, age, weight, height, BMI, albumin, Ca, phosphate, ALP, i-PTH, HbA1c, glucose, WBC, HGB, previous fracture
16	eGFR, Cre, age, weight, height, BMI, albumin, Ca, phosphate, ALP, i-PTH, HbA1c, glucose, WBC, HGB, previous fracture
32	All eligible features
Number of selected features	Method 3: Select important features from those in Methods 1 and 2 after group discussion
6	Cre, age, weight, height, albumin, HGB
7	Cre, age, weight, height, albumin, i-PTH, HGB
8	Cre, age, weight, height, albumin, i-PTH, glucose, HGB
9	Cre, age, weight, height, albumin, ALP, i-PTH, glucose, HGB
10	Cre, age, weight, height, albumin, ALP, i-PTH, glucose, WBC, HGB
15	Cre, sex, age, weight, height, albumin, Ca, phosphate, ALP, i-PTH, glucose, WBC, HGB, OA, previous fracture
16	Cre, sex, age, weight, height, albumin, Ca, phosphate, ALP, i-PTH, glucose, WBC, HGB, OA, CVA, previous fracture
32	All eligible features

**Table 2.** Feature selection methods and selected features in our study. ALP, alkaline phosphatase; BMI, body mass index; Ca, calcium; Cre, creatinine; CVA, cerebrovascular accident; eGFR, estimated glomerular filtration rate calculated using the 2021 Chronic Kidney Disease Epidemiology Collaboration creatinine equation; HbA1c, glycated hemoglobin; HGB, hemoglobin; i-PTH, intact parathyroid hormone; OA, osteoarthritis; WBC, white blood cell.

### Performance evaluation of the prediction models

Figure 2 presents the AUC results of the three feature selection methods for both the validation and test datasets. Across all three methods, we tested sets containing 32, 16, 15, 10, 9, 8, 7, and 6 features to evaluate the AUC performance of the nine ML models. In method 3, we found that selecting eight features and using the ANN model to predict osteoporosis consistently produced the best results (AUC = 0.94, validation dataset; AUC = 0.93, test dataset), as shown in both the validation dataset (Fig. 2A) and the test dataset (Fig. 2B). The eight features selected by method 3 were Cre, age, weight, height, albumin, i-PTH, glucose, and HGB (Table 2).

Table 4 presents the performance metrics for the models predicting osteoporosis using method 3 with the eight selected features. In addition to achieving the best AUC, the ANN model demonstrated the highest accuracy, precision, recall, and F1 scores for both the validation and test datasets. The CatBoost model ranked second in terms of accuracy, precision, recall, F1 score, and AUC.

Figure 3 shows the ROC curves for different ML models predicting osteoporosis, using method 3 with eight selected features, on the validation (A) and test (B) datasets. These models included ANN, CatBoost, LightGBM, XGBoost, random forest, SVM, KNN, logistic regression, and decision trees. The AUC values for each model are shown in the Figure legend. The ANN model achieved the highest AUC for both the validation and test datasets, while the CatBoost model ranked second in terms of AUC performance.

### Confusion matrix and predictive probabilities histograms

Figure 4 presents the performance of the ANN model for osteoporosis prediction using method 3 with eight selected features on the (A) validation dataset and (B) test dataset. The ANN model could distinguish the osteoporosis from the control group in both the validation data set (Fig. 4A) and the testing data set (Fig. 4B).

### Calibration curve and decision curve analysis

To plot the calibration curve and perform a decision curve analysis, we selected five models with an AUC greater than 0.85: ANN, CatBoost, LightGBM, Random Forest, and XGBoost. Figure 5 compares the performances of the five selected models in terms of the calibration curve (Fig. 5A) and decision curve analysis (Fig. 5B). Calibration and decision curve analyses further demonstrated the reliability and applicability of the ANN model.

Variable	Overall (n = 6614)		Osteoporosis group (n = 4526)		Control Group (n = 2088)		P-Value
Age (years)	75.3	(66.3–82.5)	78.0	(70.6–83.9)	67.9	(57.0–76.2)	<0.001
Female sex	4632	(70.03%)	3350	(74.02%)	1282	(61.40%)	<0.001
Height (cm)	155.1	(149.8–161.4)	153.0	(148.1–159.5)	158.4	(153.0–165.0)	<0.001
Weight (kg)	58.4	(50.8–66.0)	56.1	(49.0–63.6)	62.4	(54.2–70.0)	<0.001
BMI(kg/m <sup>2</sup> )	24.1	(21.4–26.9)	23.7	(21.0–26.6)	24.7	(21.8–27.3)	<0.001
eGFR (mL/min/1.73 m <sup>2</sup> )	47.34	(32.21–54.98)	46.21	(31.55–54.53)	49.16	(34.19–55.78)	<0.001
Serum creatinine (mg/dL)	1.33	(1.12–1.80)	1.31	(1.10–1.80)	1.38	(1.18–1.82)	<0.001
Serum albumin (g/dL)	3.9	(3.5–4.2)	3.8	(3.4–4.2)	4.1	(3.8–4.3)	<0.001
ALP (mg/dL)	95.0	(73.0–129.0)	98.0	(75.0–134.0)	89.0	(70.0–118.0)	<0.001
Corrected Ca (mg/dL)	9.0	(8.5–9.4)	8.9	(8.4–9.3)	9.1	(8.7–9.5)	<0.001
Phosphorus (mg/dL)	3.6	(3.1–4.2)	3.6	(3.1–4.1)	3.7	(3.1–4.3)	<0.001
i-PTH (pg/mL)	86.0	(46.1–211.4)	80.8	(45.3–202.2)	94.3	(48.5–232.3)	<0.001
WBC count (cells/uL)	6800	(5380–8973)	6900	(5400–9278)	6600	(5350–8500)	<0.001
Hemoglobin (g/dL)	11.2	(9.8–12.6)	11.0	(9.6–12.4)	11.7	(10.3–13.1)	<0.001
HbA1c (%)	5.9	(5.5–6.5)	5.9	(5.5–6.6)	5.9	(5.5–6.5)	0.383
Glucose (mg/dL)	107.0	(93.0–136.0)	110.0	(94.0–142.0)	103.0	(92.0–125.0)	<0.001
Previous fracture	3379	(51.09%)	2758	(60.94%)	621	(29.74%)	<0.001
OA	4309	(65.15%)	3093	(68.34%)	1216	(58.24%)	<0.001
RA	912	(13.79%)	619	(13.68%)	293	(14.03%)	0.696
DM	3018	(45.63%)	2091	(46.20%)	927	(44.40%)	0.171
HTN	4826	(72.97%)	3299	(72.89%)	1527	(73.13%)	0.837
CVA	1763	(26.66%)	1337	(29.54%)	426	(20.40%)	<0.001
Cataract	2483	(37.54%)	1764	(38.97%)	719	(34.43%)	<0.001
T1DM	82	(1.24%)	50	(1.10%)	32	(1.53%)	0.144
Hyperthyroidism	548	(8.29%)	362	(8.0%)	186	(8.9%)	0.212
Hypogonadism	73	(1.10%)	37	(0.82%)	36	(1.72%)	0.001
POI	7	(0.11%)	3	(0.07%)	4	(0.19%)	0.217
Malnutrition	90	(1.36%)	76	(1.68%)	14	(0.67%)	<0.001
CLD	2126	(32.14%)	1409	(31.13%)	717	(34.34%)	0.009
Steroid usage	5168	(78.14%)	3495	(77.22%)	1673	(80.12%)	0.008

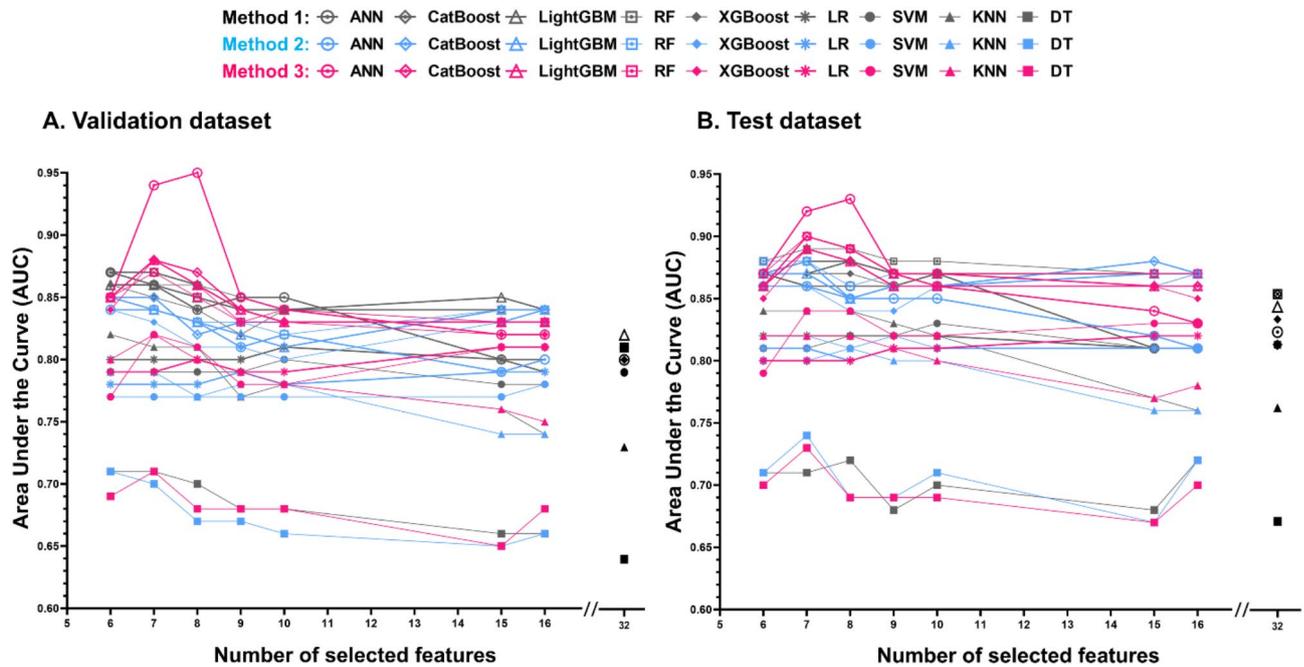
**Table 3.** Comparison of demographic and clinical characteristics of the participants. The *P*-value represents comparison between the osteoporosis and control groups. For continuous variables, the independent-samples *t* test was used if the data were normally distributed, and if not, the Mann–Whitney *U* test was used. Categorical data are represented were compared using the chi-square test or Fisher’s exact test. Values are expressed as number (percentage) or median (interquartile range). ALP, alkaline phosphatase; BMI, body mass index; Ca, calcium; CLD, chronic liver disease; CVA, cerebrovascular accident; DM, diabetes mellitus; eGFR, estimated glomerular filtration rate calculated using the 2021 Chronic Kidney Disease Epidemiology Collaboration creatinine equation; HbA1c, glycated hemoglobin; HTN, hypertension; i-PTH, intact parathyroid hormone; OA, osteoarthritis; POI, premature ovarian insufficiency; RA, rheumatoid arthritis; T1DM, type 1 diabetes mellitus; WBC, white blood cell.

The calibration curve (Fig. 5A) was used to evaluate the agreement between the predicted probabilities and actual outcomes. The XGBoost and ANN models were relatively close to the perfect calibration line, indicating strong calibration, whereas Random Forest, CatBoost, and LightGBM showed some deviations, indicating that they may slightly over- or under-predict the true probability in certain ranges.

The decision curve analysis (Fig. 5B) compared the net benefits of each model across different threshold probabilities. The ANN model exhibited the highest net benefit for most threshold probabilities, indicating that it provided the most clinically useful predictions across a wide range of decision thresholds. Other models followed closely, with varying performance across thresholds, but generally offered a net benefit above the “ALL” and “NONE” strategies. XGBoost tended to dip more steeply at higher thresholds, indicating that it may not perform as well in these cases.

### Feature importance

Figure 6 illustrates the feature importance for osteoporosis prediction, based on eight selected features selected by method 3 (age, i-PTH, weight, height, HGB, glucose, Cre, and albumin) for the ANN and CatBoost models. The SHAP Summary Plot ranks features based on the variability of SHAP values, highlighting their overall impact on the model. In contrast, the SHAP Importance Ranking Plot ranks features by the mean absolute SHAP



**Fig. 2.** Results of AUC in three different feature selection methods on (A) the validation and (B) the test datasets.

values, representing their average influence. Differences in rankings between these two plots are expected, as they arise from different calculation methods.

Age remains the most critical factor for predicting osteoporosis in both models, with higher age generally increasing the model's predicted likelihood of osteoporosis. In the ANN model, CKD-related features such as i-PTH and weight are among the top contributors. Similarly, in the CatBoost model, weight and i-PTH also play significant roles, though their relative importance varies slightly. Despite these differences, the overall ranking of feature importance is consistent between the two models, highlighting the significant role of age, weight, and i-PTH in osteoporosis prediction.

## Discussion

Our study demonstrated that the ANN model, utilizing eight selected features (age, i-PTH, weight, height, HGB, glucose, Cre, and albumin), achieved the highest predictive performance for osteoporosis in patients with stages 3–5 CKD and ESKD. The model yielded an AUC of 0.940 for the validation set and 0.930 for the test set, outperforming other ML models, including logistic regression, XGBoost, LightGBM, CatBoost, SVM, decision trees, random forests, and KNN. In addition, the ANN model exhibited strong calibration and decision curve performance, highlighting its clinical reliability and applicability. These findings suggest that ML, especially the ANN model, has significant potential for improving osteoporosis screening in patients with CKD stages 3–5 and ESKD.

Patients with stages 3–5 CKD and ESKD often experience CKD-MBD, including vascular calcification, secondary hyperparathyroidism, and renal osteodystrophy. For patients at risk for CKD-MBD or osteoporosis, BMD testing is recommended if it may influence treatment decisions<sup>7</sup>. Although DXA can overestimate bone strength and underestimate fracture risk<sup>2</sup>, particularly in advanced CKD, it remains useful because these patients have higher fracture rates than the general population, with hip fractures leading to significant morbidity and mortality rates. BMD testing is valuable if it results in interventions, such as fall prevention or osteoporosis treatment. However, many patients with CKD do not undergo BMD testing because of limited access to DXA scanning, highlighting the need for alternative clinical measures to predict osteoporosis and identify patients who should undergo DXA scanning. Kuang et al.<sup>42</sup> developed nomogram-based models to predict osteoporosis risk in 326 pre-dialysis patients categorized according to sex and renal function stage. By using multivariate logistic regression of clinical and laboratory data, these models offer a graphical tool that links patient variables to event probabilities. The nomogram models varied according to the sex and CKD stage. The AUCs for these models ranged from 0.839 to 0.933, with the best results similar to ours. One of the differences between this study and theirs is that we extended our analysis by incorporating calibration and decision curve analyses, further demonstrating the reliability and applicability of the ANN model.

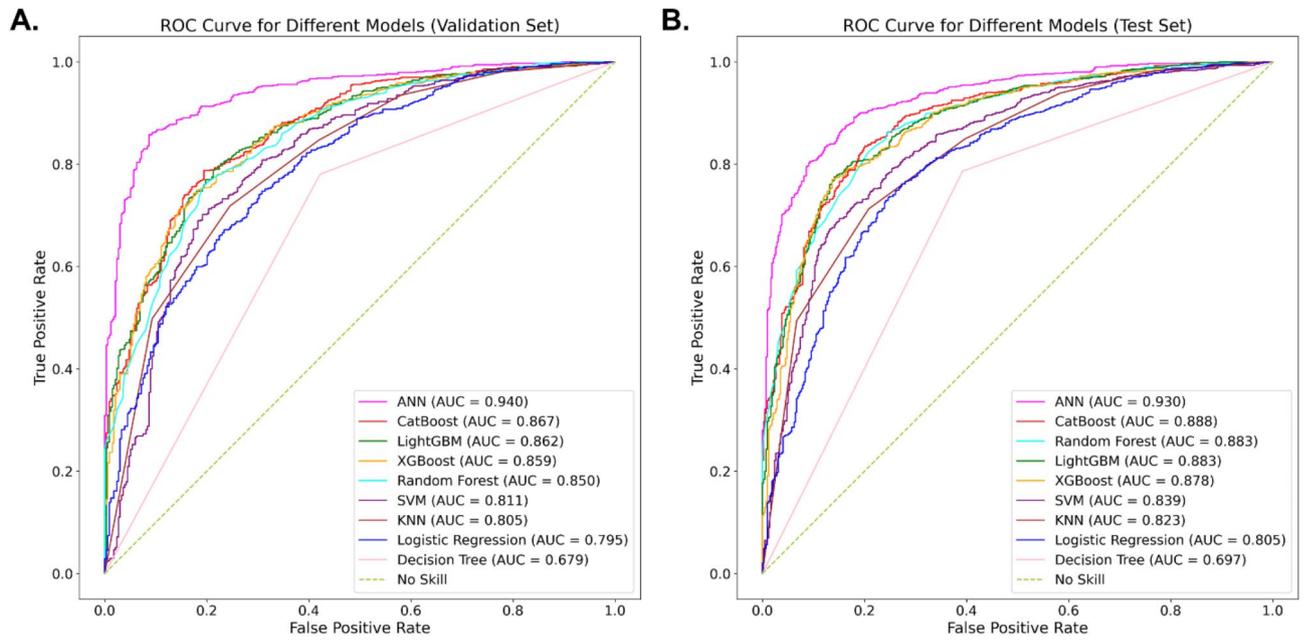
Several studies have explored the use of ML to predict osteoporosis, focusing on various populations, including the general population<sup>12–20</sup>, postmenopausal women<sup>21–24</sup>, patients with breast cancer<sup>25</sup>, patients with rheumatoid arthritis<sup>26</sup>, and those with type 2 diabetes mellitus<sup>27,28</sup>. However, to our knowledge, no previous studies have employed multiple ML models to predict osteoporosis in patients with CKD stages 3–5 and ESKD. We summarize these studies in Supplemental Tables S2 and S3, which detail the target population, sample

Models	Datasets		Accuracy	Precision	Recall	F1 Score	AUC
LR	Validation	Value	0.765	0.787	0.900	0.840	0.796
		95% CI	(0.739–0.790)	(0.758–0.814)	(0.879–0.921)	(0.820–0.860)	(0.766–0.825)
	Test	Value	0.767	0.798	0.882	0.838	0.805
		95% CI	(0.745–0.789)	(0.774–0.823)	(0.862–0.902)	(0.821–0.855)	(0.779–0.828)
XGBoost	Validation	Value	0.806	0.845	0.878	0.861	0.859
		95% CI	(0.784–0.829)	(0.818–0.870)	(0.856–0.901)	(0.843–0.879)	(0.836–0.880)
	Test	Value	0.823	0.854	0.894	0.873	0.878
		95% CI	(0.803–0.843)	(0.830–0.876)	(0.874–0.914)	(0.857–0.889)	(0.858–0.897)
LightGBM	Validation	Value	0.802	0.838	0.881	0.859	0.862
		95% CI	(0.777–0.825)	(0.811–0.864)	(0.858–0.903)	(0.839–0.878)	(0.837–0.884)
	Test	Value	0.824	0.857	0.891	0.873	0.883
		95% CI	(0.803–0.843)	(0.835–0.877)	(0.870–0.909)	(0.858–0.888)	(0.863–0.901)
CatBoost	Validation	Value	0.804	0.839	0.885	0.861	0.867
		95% CI	(0.778–0.827)	(0.810–0.863)	(0.860–0.908)	(0.841–0.879)	(0.843–0.889)
	Test	Value	0.838	0.864	0.905	0.884	0.888
		95% CI	(0.818–0.858)	(0.841–0.886)	(0.886–0.925)	(0.869–0.900)	(0.869–0.906)
SVM	Validation	Value	0.781	0.806	0.894	0.848	0.811
		95% CI	(0.755–0.806)	(0.779–0.836)	(0.872–0.914)	(0.829–0.868)	(0.781–0.840)
	Test	Value	0.785	0.817	0.883	0.849	0.839
		95% CI	(0.762–0.808)	(0.791–0.842)	(0.863–0.905)	(0.831–0.866)	(0.816–0.861)
DT	Validation	Value	0.717	0.801	0.781	0.791	0.679
		95% CI	(0.689–0.743)	(0.771–0.832)	(0.752–0.811)	(0.767–0.815)	(0.650–0.710)
	Test	Value	0.730	0.813	0.787	0.800	0.697
		95% CI	(0.707–0.754)	(0.785–0.838)	(0.761–0.811)	(0.780–0.819)	(0.670–0.724)
RF	Validation	Value	0.798	0.827	0.891	0.858	0.850
		95% CI	(0.772–0.822)	(0.799–0.854)	(0.867–0.913)	(0.838–0.877)	(0.825–0.874)
	Test	Value	0.830	0.859	0.899	0.879	0.883
		95% CI	(0.810–0.851)	(0.837–0.883)	(0.879–0.918)	(0.864–0.894)	(0.863–0.902)
KNN	Validation	Value	0.763	0.814	0.846	0.830	0.804
		95% CI	(0.737–0.788)	(0.784–0.843)	(0.819–0.873)	(0.808–0.850)	(0.776–0.829)
	Test	Value	0.772	0.822	0.849	0.836	0.823
		95% CI	(0.751–0.795)	(0.799–0.847)	(0.827–0.872)	(0.818–0.854)	(0.801–0.845)
ANN	Validation	Value	<b>0.873*</b>	<b>0.897*</b>	<b>0.920*</b>	<b>0.909*</b>	<b>0.940*</b>
		95% CI	(0.855–0.891)	(0.876–0.917)	(0.901–0.939)	(0.894–0.922)	(0.925–0.954)
	Test	Value	<b>0.861*</b>	<b>0.885*</b>	<b>0.916*</b>	<b>0.900*</b>	<b>0.930*</b>
		95% CI	(0.841–0.878)	(0.863–0.904)	(0.897–0.934)	(0.885–0.914)	(0.916–0.942)

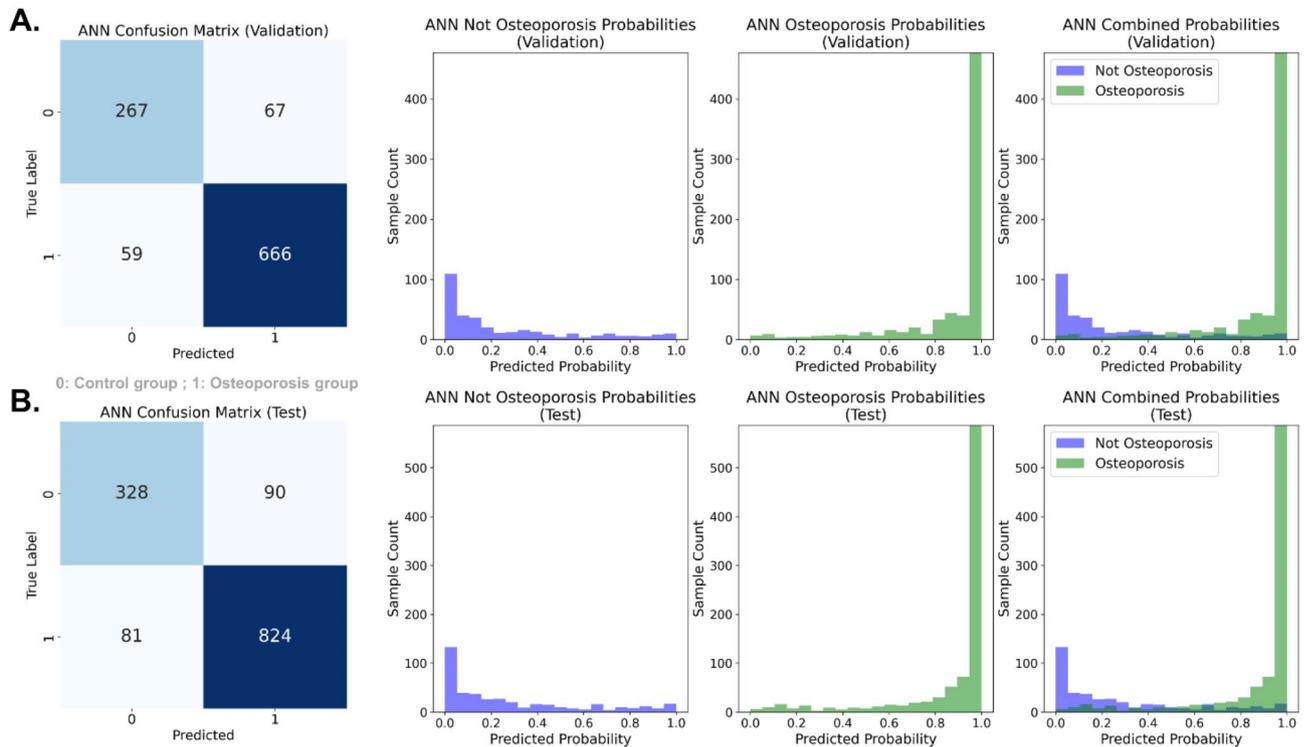
**Table 4.** Performance metrics for nine ML models in the method 3 with eight selected features. \*The best result is indicated in bold. ANN, artificial neural network; AUC, area under the receiver operating characteristic curve; CatBoost, category boosting; CI, confidence interval; DT, decision tree; KNN, k-nearest neighbors algorithm; LightGBM, light gradient boosting machine; LR, logistic regression; RF, random forest; SVM, support vector machine; XGBoost, extreme gradient boosting.

size, feature selection methods, number of selected features, best-performing models, and AUC scores for osteoporosis prediction across different populations, such as postmenopausal women and patients with breast cancer. The AUC of these studies ranged from 0.710 to 0.961. In comparison, our study, which focused on CKD stages 3–5 and ESKD, achieved an AUC of 0.930 using the ANN model, which was higher than those of many previous studies. A likely reason for this is our targeted selection of key CKD-related features such as parathyroid hormone levels, which may not be significant in non-CKD populations. In contrast, studies by Inui et al.<sup>23</sup> (AUC of 0.961 using LightGBM in elderly women) and Kwon et al.<sup>24</sup> (AUC of 0.921 using AdaBoost in postmenopausal women) reported competitive results but targeted different populations and used different methodologies.

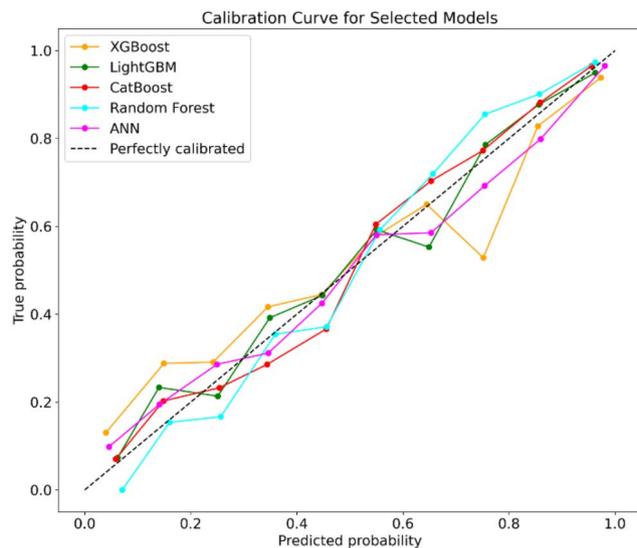
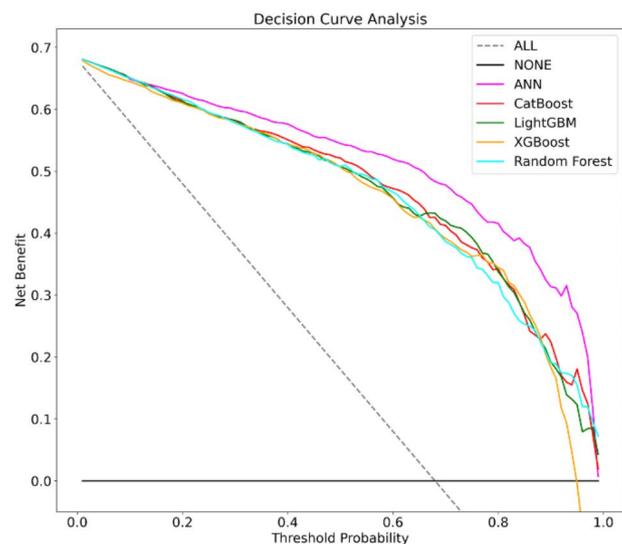
Our study enhances the understanding of osteoporosis prediction in patients with CKD by emphasizing the importance of CKD-related features such as intact parathyroid hormone, which has not been as prominently highlighted in previous models for the general population. These findings suggest that tailored models for specific patient populations, such as patients with CKD, can substantially improve predictive accuracy. Furthermore, the strong performance of our ANN model emphasizes the potential of ML to guide clinical decision-making, offering a noninvasive and cost-effective method to identify patients at high risk of osteoporosis in CKD populations. This could optimize the allocation of resources for bone density scans and improve patient outcomes.



**Fig. 3.** ROC curve comparison for different models on (A) the validation and (B) the test datasets. The y-axis represents the true-positive rate, whereas the x-axis represents the false-positive rate. The dashed green line represents the no-skilled classifier.



**Fig. 4.** Confusion matrices and predictive probabilities histograms for the ANN model on (A) the validation and (B) the test datasets. The confusion matrices on the left indicate the number of correct and incorrect classifications, where 0 represents the control group (no osteoporosis) and 1 represents the osteoporosis group. The middle panels display histograms of the predicted probabilities for the “not osteoporosis” and “osteoporosis” classes separately. The rightmost panels combine the predicted probabilities of both classes and show the overall distribution. The x-axis represents the predicted probability, and the y-axis represents the sample count, reflecting the confidence of the model in its predictions.

**A. Calibration Curve****B. Decision Curve Analysis**

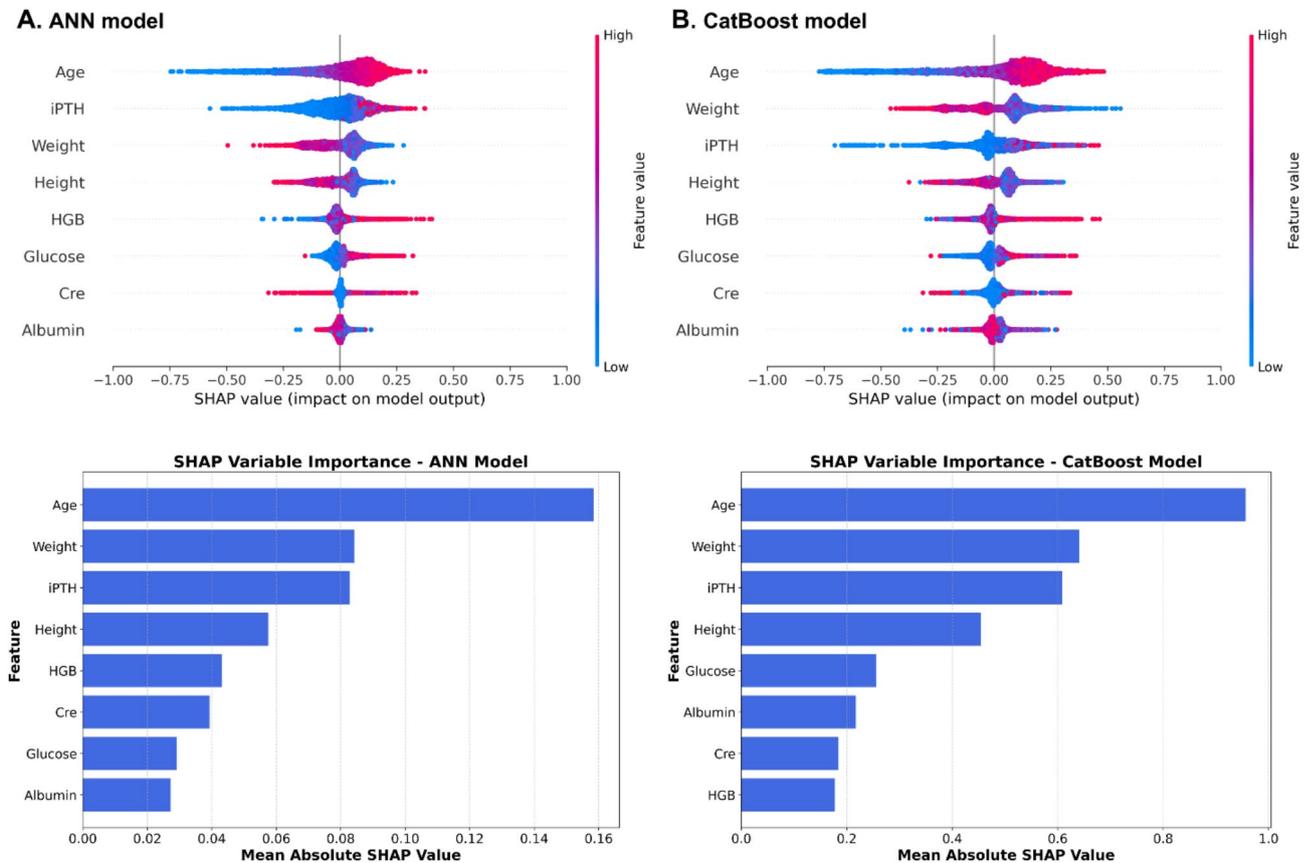
**Fig. 5.** Calibration curve (A) and decision curve analysis (B) for five selected models in osteoporosis prediction. For the calibration curve, the perfectly calibrated model is represented by the dashed diagonal. Deviations from this line indicate calibration issues, with models that either overestimate or underestimate the probabilities. For the decision curve analysis, the dashed gray line represents the “ALL” strategy, which assumes that every patient has osteoporosis, while the black line represents the “NONE” strategy, which assumes that no patient has osteoporosis. The x-axis represents the threshold probability, and the y-axis represents the net benefit. The models closer to the top of the graph provide better clinical utility across different threshold probabilities.

In this study, we addressed the category imbalance between the osteoporosis group (4526 patients) and the control group (2088 patients) using several strategies to enhance model robustness and minimize bias. First, we conducted a pilot study using three data imbalance handling techniques (Downsampling, SMOTE, and Tomek Links) on the entire dataset. The impact of these techniques on the performance of nine different models is summarized in Supplemental Table S4. However, none of these techniques significantly improved model performance compared to the original dataset. Subsequently, we employed stratified random sampling during dataset splitting to maintain the original class distribution across the training, validation, and test sets, ensuring adequate representation of the control group. To comprehensively evaluate model performance, we used multiple metrics beyond accuracy, including F1-score, AUC and recall. Furthermore, we utilized SHAP values to analyze feature importance, confirming that predictions were driven by clinically relevant variables rather than artifacts of imbalance. These strategies collectively ensured that our models effectively predicted osteoporosis in patients with CKD stages 3–5 and ESKD.

This study has several limitations. First, the retrospective design may have introduced causal inferences and selection bias. Notably, the prevalence of osteoporosis in our cohort (68.4%) was significantly higher than the 18–32% reported in other CKD studies. This discrepancy is likely due to Taiwan’s National Health Insurance policies, which restrict DXA screening to patients with CKD stages 3–5 and those with ESKD who are at higher risk for osteoporosis. Consequently, our study population may not fully represent the general CKD population, potentially limiting the generalizability of our findings. This selection bias is further reflected in Table 3, where 51.09% of participants had a history of fractures. Future research with a larger CKD population and less restricted screening criteria are required to validate our findings.

Second, this study used missing data imputation to handle incomplete variables. Among the eight selected features, three (height, weight, and i-PTH) had high missing rates (height: 34.59%, weight: 30.95%, i-PTH: 57.51%). To ensure the robustness of our results, we conducted a sensitivity analysis using various missing data handling methods, including multiple imputation (iterative imputer), mean imputation, KNN imputation, and complete-case analysis (drop missing). The results demonstrated that our model’s predictive performance remained consistent across different imputation strategies (Supplemental Table S5). While multiple imputation yielded fair performance, the overall findings were not significantly affected by the choice of imputation method, confirming the robustness of our predictive model. However, the high percentage of missing values in key variables remains a limitation, and future studies should aim to collect more complete datasets for further validation.

Third, a cross-sectional survey was employed in the study, which may not have been entirely representative, as it only provided a moment-in-time view of the population. Fourth, only internal validation was performed; external validation using different datasets was necessary. Fifth, the study had a relatively small sample size and was conducted at a single hospital, highlighting the need for larger multicenter multinational studies to confirm the robustness of our model. Because the selected features are commonly measured in patients



**Fig. 6.** SHAP summary plot and variable importance ranking plot for (A) ANN and (B) CatBoost models using method 3 with eight selected features. The upper panels display the SHAP summary plots, where the x-axis represents the SHAP values, reflecting the impact of each feature on the model output. The color scale represents the feature values, with blue indicating lower values and pink indicating higher values. A positive SHAP value indicates that the feature increases the likelihood of predicting osteoporosis, whereas a negative SHAP value reduces this likelihood. The lower panels present the variable importance ranking plots, showing the mean absolute SHAP values for each feature, which represent their overall contribution to the model's predictions.

with CKD stages 3–5 and ESKD in Taiwan, external validation using data from other hospitals could be considered. Additionally, the majority of participants were Taiwanese, and osteoporosis risk factors may vary across different ethnic groups and environments. Therefore, the applicability of our machine-learning models should be tested in diverse populations worldwide. These limitations may have affected the generalizability of the prediction model. Future research should address these issues by integrating more data sources or developing advanced preprocessing techniques to manage biased data effectively.

In conclusion, we aimed to develop ML models to predict osteoporosis in patients with CKD stages 3–5 and ESKD. By applying various models, we found that the ANN model, using eight selected features (age, creatinine, height, weight, albumin, glucose, intact parathyroid hormone, and hemoglobin), achieved the highest predictive performance with an AUC of 0.940 on the validation dataset and 0.930 on the test dataset. These results suggest that the ANN model has the potential for clinical application in screening high-risk patients with CKD and ESKD for osteoporosis. Our findings successfully address this research question by providing a predictive tool that can help optimize resource allocation for BMD scanning in clinical settings. Although our study is limited by its retrospective nature, the strong performance of the ANN model provides a foundation for future research on predicting osteoporosis, particularly in populations with CKD. Future studies should focus on external validation using larger and more diverse datasets to enhance the generalizability of the model. The implications of this research suggest that ML, particularly ANN models, can improve osteoporosis screening in patients with CKD. We hypothesized that this would further enhance the management of osteoporosis in these patients, ultimately improving outcomes and reducing healthcare costs.

### Data availability

All data generated or analyzed during this study are included in this article. The data and code have been placed in public GitHub and are available at <https://github.com/jaten000/OsteoporosisPredictionInCKD>.

Received: 4 February 2025; Accepted: 25 March 2025

Published online: 03 April 2025

## References

- Zaimi, M. & Grapsa, E. Current therapeutic approach of chronic kidney disease-mineral and bone disorder. *Ther. Apher. Dial.* **28**, 671–689 (2024).
- Abdalbary, M. et al. Management of osteoporosis in patients with chronic kidney disease. *Osteoporos. Int.* **33**, 2259–2274 (2022).
- Nickolas, T. L., McMahon, D. J. & Shane, E. Relationship between moderate to severe kidney disease and hip fracture in the United States. *J. Am. Soc. Nephrol.* **17**, 3223–3232 (2006).
- Nickolas, T. L., Leonard, M. B. & Shane, E. Chronic kidney disease and bone fracture: a growing concern. *Kidney Int.* **74**, 721–731 (2008).
- Alem, A. M. et al. Increased risk of hip fracture among patients with end-stage renal disease. *Kidney Int.* **58**, 396–399 (2000).
- Pimentel, A., Ureña-Torres, P., Zillikens, M. C., Bover, J. & Cohen-Solal, M. Fractures in patients with CKD—diagnosis, treatment, and prevention: A review by members of the European Calcified Tissue Society and the European Renal Association of Nephrology Dialysis and Transplantation. *Kidney Int.* **92**, 1343–1355 (2017).
- Kidney Disease: Improving Global Outcomes (KDIGO) CKD-MBD Update Work Group. KDIGO 2017 Clinical Practice Guideline Update for the Diagnosis, Evaluation, Prevention, and Treatment of Chronic Kidney Disease-Mineral and Bone Disorder (CKD-MBD). *Kidney Int. Suppl.* **7**, 1–59 (2017).
- Kanis, J. A. Assessment of fracture risk and its application to screening for postmenopausal osteoporosis: synopsis of a WHO report. WHO Study Group. *Osteoporos. Int.* **4**, 368–381 (1994).
- Tai, T. W. et al. (2023) Clinical practice guidelines for the prevention and treatment of osteoporosis in Taiwan: 2022 Update. *J. Formos. Med. Assoc.* **122**(1), S4–S13 (2023).
- Jiang, F. et al. Artificial intelligence in healthcare: Past, present and future. *Stroke Vasc. Neurol.* **2**, 230–243 (2017).
- Yuan, Q. et al. Role of artificial intelligence in kidney disease. *Int. J. Med. Sci.* **17**, 970–984 (2020).
- Erjiang, E. et al. Machine learning can improve clinical detection of low BMD: The DXA-HIP study. *J. Clin. Densitom.* **24**, 527–537. <https://doi.org/10.1016/j.jocd.2020.10.004> (2021).
- 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.* **18**, 7635 (2021).
- Wang, Y. et al. Prediction model for the risk of osteoporosis incorporating factors of disease history and living habits in physical examination of population in Chongqing, Southwest China: Based on artificial neural network. *BMC Public Health* **21**, 991 (2021).
- Iliou, T., Anagnostopoulos, C. N. & Anastassopoulos, G. Osteoporosis detection using machine learning techniques and feature selection. *Int. J. Artif. Intell. Tools.* **23**, 1450014 (2014).
- Park, H. W. et al. Application of machine learning to identify clinically meaningful risk group for osteoporosis in individuals under the recommended age for dual-energy x-ray absorptiometry. *Calcif. Tissue Int.* **109**, 645–655 (2021).
- Qiu, C. et al. Developing and comparing deep learning and machine learning algorithms for osteoporosis risk prediction. *Front. Artif. Intell.* **7**, 1355287 (2024).
- Wu, X. & Park, S. A prediction model for osteoporosis risk using a machine-learning approach and its validation in a large cohort. *J. Korean Med. Sci.* **38**, e162 (2023).
- Albuquerque, G. A. et al. Osteoporosis screening using machine learning and electromagnetic waves. *Sci. Rep.* **13**, 12865 (2023).
- Tu, J. B., Liao, W. J., Liu, W. C. & Gao, X. H. Using machine learning techniques to predict the risk of osteoporosis based on nationwide chronic disease data. *Sci. Rep.* **14**, 5245 (2024).
- Shim, J. G. et al. Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women. *Arch. Osteoporos.* **15**, 169 (2020).
- Bui, H. M. et al. Predicting the risk of osteoporosis in older Vietnamese women using machine learning approaches. *Sci. Rep.* **12**, 20160 (2022).
- Inui, A. et al. Screening for osteoporosis from blood test data in elderly women using a machine learning approach. *Bioengineering* **10**, 277 (2023).
- Kwon, Y. et al. Osteoporosis pre-screening using ensemble machine learning in postmenopausal Korean women. *Healthcare* **10**, 1107 (2022).
- Ji, L. et al. Osteoporosis, fracture and survival: Application of machine learning in breast cancer prediction models. *Front. Oncol.* **12**, 973307 (2022).
- Lee, C., Joo, G., Shin, S., Im, H. & Moon, K. W. Prediction of osteoporosis in patients with rheumatoid arthritis using machine learning. *Sci. Rep.* **13**, 21800 (2023).
- Wu, X. et al. Development of machine learning models for predicting osteoporosis in patients with type 2 diabetes mellitus—A preliminary study. *Diabetes Metab. Syndr. Obes.* **16**, 1987–2003 (2023).
- Wu, X. et al. Application of machine learning algorithms to predict osteoporosis in postmenopausal women with type 2 diabetes mellitus. *J. Endocrinol. Invest.* **46**, 2535–2546 (2023).
- Wu, C. H. et al. Consensus of official position of IOF/ISCD FRAX initiatives in Asia-Pacific region. *J. Clin. Densitom.* **17**, 150–155 (2014).
- Fouque, D. et al. A proposed nomenclature and diagnostic criteria for protein-energy wasting in acute and chronic kidney disease. *Kidney Int.* **73**, 391–398 (2008).
- Inker, L. A. et al. New creatinine- and cystatin C-based equations to estimate GFR without race. *N. Engl. J. Med.* **385**, 1737–1749 (2021).
- Kramer, H. J. et al. An endorsement of the removal of race from GFR estimation equations: A position statement from the National Kidney Foundation kidney disease outcomes quality initiative. *Am. J. Kidney Dis.* **80**, 691–696 (2022).
- Payne, R. B., Little, A. J., Williams, R. B. & Milner, J. R. Interpretation of serum calcium in patients with abnormal serum proteins. *Br. Med. J.* **4**, 643–646 (1973).
- Zhao, J., Lui, H., Kalia, S., Lee, T. K. & Zeng, H. Improving skin cancer detection by Raman spectroscopy using convolutional neural networks and data augmentation. *Front. Oncol.* **14**, 1320220 (2024).
- Hsu, C. T., Pai, K. C., Chen, L. C., Lin, S. H. & Wu, M. J. Machine learning models to predict the risk of rapidly progressive kidney disease and the need for nephrology referral in adult patients with type 2 diabetes. *Int. J. Environ. Res. Public Health.* **20**, 3396 (2023).
- Li, C. Little's test of missing completely at random. *Stata J.* **13**, 795–809 (2013).
- Emmanuel, T. et al. A survey on missing data in machine learning. *J. Big Data.* **8**, 140 (2021).
- Schmitt, S. & Rothlauf, F. Comparison of imputation methods for categorical real-world prostate cancer data with natural order. *Stud. Health Technol. Inform.* **316**, 1800–1804 (2024).
- Pan, Y. et al. Cognitive frailty in relation to vitamin B12 and 25-hydroxyvitamin D in an elderly population: A cross-sectional study from NHANES. *Front. Nutr.* **11**, 1430722 (2024).
- Tseng, P. Y. et al. Prediction of the development of acute kidney injury following cardiac surgery by machine learning. *Crit. Care.* **24**, 478 (2020).
- Pedregosa, F. et al. Scikit-learn: machine learning in Python. *J. Mach. Learn. Res.* **12**, 2825–2830 (2011).

42. Kuang, C. et al. Risk factors and clinical prediction models for osteoporosis in pre-dialysis chronic kidney disease patients. *Ren. Fail.* **46**, 2361802 (2024).

### Acknowledgements

The authors thank the Osteoporosis Prevention and Treatment Center of the Taichung Veterans General Hospital for providing the osteoporosis database.

### Author contributions

CT. H., CH. C., SY. L., MJ. W. and CY. H. made substantial contributions to the conception and design of the study and interpretation of data. CT. H., YL. D. and SY. L. gathered, analyzed, and interpreted the data. CT. H. wrote the original draft of the manuscript. CH. C., MJ. W., and CT. H. supervised the data interpretation and substantively revised the manuscript. All the authors have read and approved the final version of the manuscript.

### Declarations

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-95928-5>.

**Correspondence** and requests for materials should be addressed to C.-Y.H.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

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

© The Author(s) 2025