



Artificial Intelligence-enabled Chest X-ray Classifies Osteoporosis and Identifies Mortality Risk

Dung-Jang Tsai^{1,2,3} · Chin Lin^{2,3,4} · Chin-Sheng Lin⁵ · Chia-Cheng Lee^{6,7} · Chih-Hung Wang^{8,9} · Wen-Hui Fang^{3,10}

Received: 27 August 2023 / Accepted: 26 December 2023

© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2024

Abstract

A deep learning model was developed to identify osteoporosis from chest X-ray (CXR) features with high accuracy in internal and external validation. It has significant prognostic implications, identifying individuals at higher risk of all-cause mortality. This Artificial Intelligence (AI)-enabled CXR strategy may function as an early detection screening tool for osteoporosis. The aim of this study was to develop a deep learning model (DLM) to identify osteoporosis via CXR features and investigate the performance and clinical implications. This study collected 48,353 CXRs with the corresponding T score according to Dual energy X-ray Absorptiometry (DXA) from the academic medical center. Among these, 35,633 CXRs were used to identify CXR- Osteoporosis (CXR-OP). Another 12,720 CXRs were used to validate the performance, which was evaluated by the area under the receiver operating characteristic curve (AUC). Furthermore, CXR-OP was tested to assess the long-term risks of mortality, which were evaluated by Kaplan–Meier survival analysis and the Cox proportional hazards model. The DLM utilizing CXR achieved AUCs of 0.930 and 0.892 during internal and external validation, respectively. The group that underwent DXA with CXR-OP had a higher risk of all-cause mortality (hazard ratio [HR] 2.59, 95% CI: 1.83–3.67), and those classified as CXR-OP in the group without DXA also had higher all-cause mortality (HR: 1.67, 95% CI: 1.61–1.72) in the internal validation set. The external validation set produced similar results. Our DLM uses CXRs for early detection of osteoporosis, aiding physicians to identify those at risk. It has significant prognostic implications, improving life quality and reducing mortality. AI-enabled CXR strategy may serve as a screening tool.

Keywords Artificial intelligence · Bone mass density · T score · Chest X-ray · Deep learning

✉ Wen-Hui Fang
rumaf.fang@gmail.com

¹ Department of Statistics and Information Science, Fu Jen Catholic University, New Taipei City, Taiwan, R.O.C.

² Medical Technology Education Center, School of Medicine, National Defense Medical Center, Taipei, Taiwan, R.O.C.

³ Artificial Intelligence of Things Center, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

⁴ School of Public Health, National Defense Medical Center, Taipei, Taiwan, R.O.C.

⁵ Division of Cardiology, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

⁶ Medical Informatics Office, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

⁷ Division of Colorectal Surgery, Department of Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

⁸ Department of Otolaryngology-Head and Neck Surgery, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

⁹ Graduate Institute of Medical Sciences, National Defense Medical Center, Taipei, Taiwan, R.O.C.

¹⁰ Department of Family and Community Medicine, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

Introduction

Osteoporosis is a prevalent condition, particularly among postmenopausal women, but it often goes unnoticed until a fracture occurs. Timely detection of osteoporosis is crucial for preventing osteoporotic fractures. In the United States, the incidence of fractures related to osteoporosis is more than four times higher than that of stroke, heart attack, and breast cancer combined [1]. According to the World Health Organization's meeting report, osteoporotic fractures result in more hospital bed-days than these diseases in several high-income countries [2]. Hip fractures, which are among the most common osteoporotic fractures, can cause difficulties with walking, chronic pain, disability, loss of independence, and reduced quality of life. Shockingly, 21% to 30% of patients who suffer from hip fractures pass away within one year [3]. Based on 2009 data from Taiwan, approximately 16,000 individuals experience hip fractures annually, with women being twice as likely as men to be affected. Furthermore, the incidence of hip fractures increases significantly with age, with Taiwanese women between the ages of 70 and 80 having a 10% chance of experiencing a hip fracture [4]. The heightened potential for fragility fractures increases the likelihood of elevated risks in terms of all-cause mortality, as well as mortality attributed to cardiovascular diseases (CVDs) and cancer [5–7]. In a US study, osteoporosis increased all-cause mortality risk in total femur, femur neck, and intertrochanter areas, but not significantly for cancer or cardiovascular (CV) mortality [8].

Currently, the most reliable method of diagnosing osteoporosis is to measure bone mineral density (BMD) in the hip and lumbar spine using Dual energy X-ray Absorptiometry (DXA) [9]. According to the guidelines established by the World Health Organization (WHO), a BMD measurement that falls at or below 2.5 standard deviations from the young adult mean ($T \text{ score} \leq -2.5$) indicates osteoporosis, while a T score ranging between -1.0 and -2.5 at any location indicates low bone mass or osteopenia. In addition, the US Preventive Services Task Force recommends BMD testing for women aged 65 and above as a preventive measure against osteoporotic fractures [3]. Despite its effectiveness, DXA has some disadvantages, such as the high cost of equipment and the risk of radiation exposure [10, 11]. Raising awareness about osteoporosis may be the most effective approach to prevent osteoporotic fractures [12]. Unfortunately, elderly individuals have a low level of awareness regarding this disease [13]. Nevertheless, the present screening rates are notably insufficient [14], as only a minor proportion of eligible individuals undergo DXA examination [15]. This underscores the necessity for a risk assessment tool that promotes screening within high-risk populations [16].

A promising strategy for identifying individuals at risk of osteoporosis and fragility fractures is opportunistic screening through imaging methods other than DXA. This approach involves using radiographs that have already been taken for other clinical purposes, with no additional cost, time, or radiation exposure to the patient. For instance, several studies have utilized computed tomography (CT)-based metrics to estimate BMD [17], classify osteoporosis [18], simulate DXA T scores [19], and predict fracture risk [20]. Compared to other imaging modalities, X-ray radiography is more widely available, has broader applications, incurs lower radiation exposure and is generally more cost-effective. Furthermore, radiographs provide excellent spatial resolution, allowing for the visualization of fine bone texture that is closely associated with bone density [21]. This makes it possible to differentiate individuals with osteoporotic fractures from those without. Deep learning algorithms have surpassed traditional methods in terms of visual recognition accuracy [22], which is essential for clinical applications such as fracture detection [23, 24], retinopathy grading [25], and lung nodule identification [26]. Recent advancements in orthopedic research have paved the way for the application of DLMs in osteoporosis screening [27]. Previous studies have shown the feasibility of diagnosing osteoporosis based on radiographs of the lumbar spine and hip joint [28, 29], as well as measuring the BMD (g/cm^2) of these sites from radiographs [9, 30]. Furthermore, two studies have utilized chest X-rays (CXR) for diagnosing osteoporosis [31, 32].

Our hypothesis is that a DLM can accurately classify osteoporosis by analyzing chest radiography and uncover a heightened risk of mortality in the Artificial Intelligence (AI)-predicted positive CXR- osteoporosis (CXR-OP) group. Our model focuses on classifying individuals into categories based on T scores (normal, osteopenia, and osteoporosis) using CXRs. Given the association between osteoporosis and increased mortality risk, individuals predicted as part of the CXR-OP group in the absence of DXA examinations should exhibit potential mortality risks. We trained our model using a large dataset from a medical center, aiming to achieve excellent predictive performance for osteoporosis classification, ultimately envisioning CXRs as a screening tool for osteoporosis.

Methods

Data source and population

The institutional ethics committee of the Tri-Service General Hospital (C202105049) reviewed and approved this study, and we retrospectively developed and evaluated a DLM internally and externally. The CXRs were collected

from two hospitals, an academic medical center in Neihu District (hospital A) and a community hospital in Zhongzheng District (hospital B), from January 1, 2010, to April 30, 2021. Patients aged less than 20 years old were excluded.

Figure 1 shows the assignment of samples in this study. We identified CXRs in the posterior-anterior view with at least 1 DXA obtained within a window of ± 180 days of the index CXR. There were 24,448 patients with at least one CXR in hospital A. The 12,221 patients were used in the development set, which included 25,574 CXRs for DLM training. A total of 4,878 patients were assigned to the tuning set. These patients provided 10,059 CXRs for guiding the training process and determining the optimal operating point for subsequent use. Finally, 7,349 patients were assigned to an internal validation set, which contained only the first CXR that was used for the accuracy test and follow-up analysis. We also collected 5,371 patients in hospital B using the same inclusion criteria as the internal validation set to verify the extrapolation of the DLM.

Data collection

The BMD data (g/cm^2) were collected using a DXA scanner (Lunar Prodigy Series; GE Healthcare, Madison, WI, USA). The BMD values of the lumbar spine (anteroposterior projection at L1-L4) and the femurs (i.e., femoral neck, trochanter, and total hip) were measured. BMD and T scores for each lumbar vertebra and femur were then

calculated using Encore V13.6 software. Participants were classified into three categories based on the WHO T score classification [2]. Osteoporosis is defined as a T score of less than -2.5 ; osteopenia is defined as a T score of -1 to -2.5 ; and healthy is defined as a T score above -1 . Serial scans for each participant were performed on the same day, and reports were confirmed and judged by experienced radiologists.

The disease histories were based on the corresponding International Classification of Diseases (ICD), Ninth Revision and Tenth Revision (ICD-9 and ICD-10, respectively) described previously [33]. The primary outcome of this study was the prediction of the DXA T score. Electronic medical records defined the status (the T score value) of the patient. These records were updated by hospital staff as needed. We also performed a secondary analysis on all-cause mortality and extensive CVDs, such as CV mortality, new-onset acute myocardial infarction (AMI), new-onset stroke (STK), new-onset coronary artery disease (CAD), new-onset atrial fibrillation (Afib), and new-onset heart failure (HF). We defined a new-onset event as a record of the corresponding ICD codes, such as AMI, STK, CAD, Afib or HF. For mortality, the survival time was calculated with reference to the date of the CXR record and we only included patients with follow-up hospital visits. Mortality event was captured through the electronic medical record. Moreover, data for alive visits were censored at the patient's last known hospital alive encounter to limit bias from incomplete records.

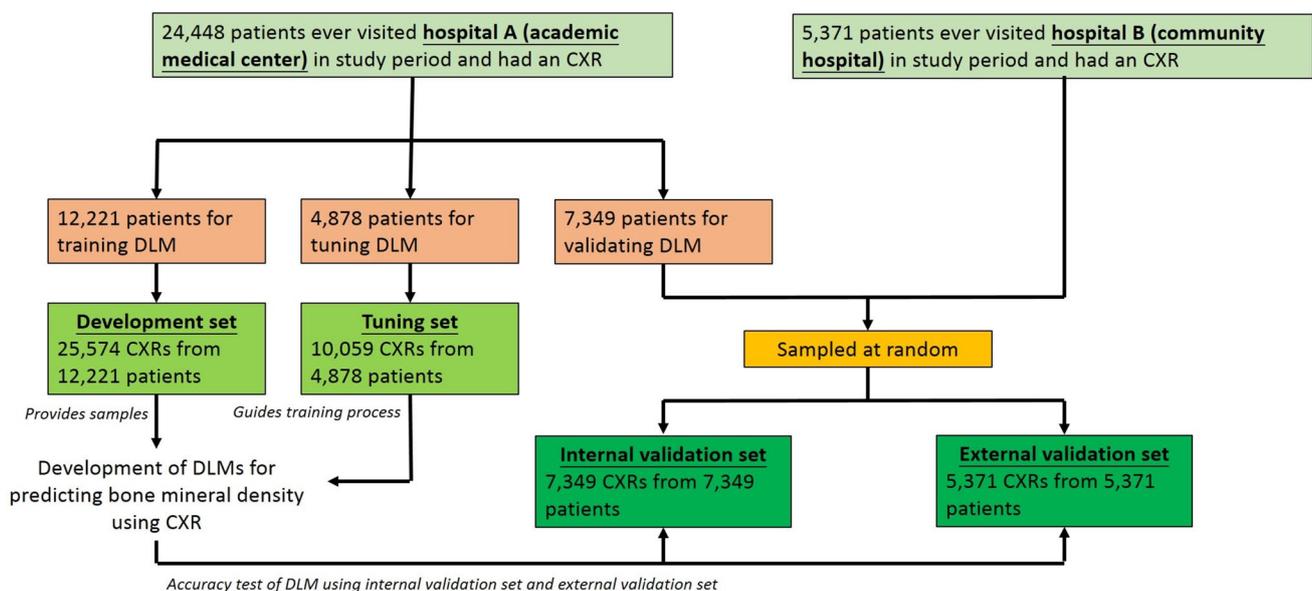


Fig. 1 Development, tuning, internal validation, and external validation set generation and CXR labeling of bone mineral density. Schematic of the dataset creation and analysis strategy, which was devised to assure a robust and reliable dataset for training, validating, and testing of the network. Once a patient's data were placed in one of

the datasets, that individual's data were used only in that set, avoiding 'cross-contamination' among the training, validation, and test datasets. The details of the flow chart and how each of the datasets was used are described in the Methods

Patients meeting any of the above criteria before the index date of the CXR were excluded and defined as having a corresponding disease history.

Implementation of the deep learning model

The CXR images were recorded in Digital Imaging and Communications in Medicine (DICOM) format with a resolution of more than 3000×3000 pixels. The training details of the DLM for CXR were revised from a previous study [34], which was a 121-layer DenseNet. We resized our CXR to let the short side be 256 pixels without changing the aspect ratio. In the training stage, we randomly cropped 224×224 pixels as input and applied a random lateral inversion with 50% probability. In the inference stage, the 10-crop evaluation was used to generate 10 probabilities for each CXR, and the final prediction was based on their average.

We trained these DLMs with a batch size of 32 and an initial learning rate of 0.001 using the Adam optimizer with standard parameters ($\beta_1 = 0.9$ and $\beta_2 = 0.999$). An oversampling process

was implemented to ensure that osteoporosis was adequately recognized. For each batch, we sampled 16 cases and 16 controls in the development set. The learning rate was decayed by a factor of 10 each time the loss on the tuning set plateaued after an epoch. To prevent the networks from overfitting, early stopping was performed by saving the network after every epoch and choosing the saved DLMs with the lowest loss on the tuning set. L2 regularization with a coefficient of 10^{-4} was also applied to avoid overfitting.

Statistical analysis

The characteristics of the different sets are presented as the means and standard deviations, numbers of patients, or percentages. Continuous variables were assessed using Student's t-test or Analysis of Variance and presented as the mean \pm SD. Categorical variables were examined using the χ^2 test or Fisher's exact test, as deemed appropriate. The performance of the DLMs was evaluated by the receiver operating characteristic (ROC) curve for implant pacemaker analysis, and the area

Table 1 | Baseline characteristics

	Development set	Tuning set	Internal validation set	External validation set	<i>p</i> value
Demographics					
GENDER					<0.001
Female	15651(61.6%)	6473(64.7%)	4430(60.9%)	3876(72.2%)	
Male	9777(38.4%)	3534(35.3%)	2845(39.1%)	1495(27.8%)	
Age (years)	62.20 \pm 16.22	62.64 \pm 16.41	57.20 \pm 15.74	62.83 \pm 14.82	<0.001
Height (cm)	159.48 \pm 8.73	158.87 \pm 8.50	159.86 \pm 8.92	159.75 \pm 8.46	<0.001
Weight (kg)	59.83 \pm 12.25	59.96 \pm 12.69	61.27 \pm 12.46	61.37 \pm 12.50	<0.001
BMI (kg/m ²)	23.72 \pm 3.99	23.89 \pm 4.16	24.09 \pm 3.77	24.06 \pm 4.01	<0.001
BMD	-0.95 \pm 1.74	-0.96 \pm 1.74	-0.53 \pm 1.75	-1.02 \pm 1.64	<0.001
BMD_Group					<0.001
Normal	12590(49.2%)	4940(49.1%)	4420(60.1%)	2547(47.4%)	
Osteopenia	7476(29.2%)	2934(29.2%)	1787(24.3%)	1753(32.6%)	
Osteoporosis	5508(21.5%)	2185(21.7%)	1142(15.5%)	1071(19.9%)	
Disease history					
AMI	281(1.1%)	100(1.0%)	34(0.5%)	37(0.7%)	<0.001
STK	3026(11.9%)	1232(12.3%)	420(5.8%)	632(11.8%)	<0.001
CAD	4063(16.0%)	1629(16.3%)	675(9.3%)	1124(20.9%)	<0.001
Afib	1110(4.4%)	508(5.1%)	112(1.5%)	151(2.8%)	<0.001
HF	2056(8.1%)	911(9.1%)	168(2.3%)	338(6.3%)	<0.001
DM	4817(18.9%)	2147(21.5%)	779(10.7%)	1167(21.7%)	<0.001
HTN	1369(5.4%)	576(5.8%)	149(2.0%)	243(4.5%)	<0.001
CKD	3985(15.7%)	1653(16.5%)	342(4.7%)	448(8.3%)	<0.001
HLP	7530(29.6%)	3113(31.1%)	1480(20.3%)	2232(41.6%)	<0.001
COPD	4434(17.4%)	1546(15.4%)	594(8.2%)	1211(22.5%)	<0.001
All-cause mortality, n (%)	3074(12.1%)	1321(13.2%)	273(3.8%)	337(6.3%)	<0.001
Follow-up (days)	1063.92 \pm 1064.57	1074.36 \pm 1055.86	1137.49 \pm 1143.11	1704.79 \pm 1164.90	<0.001

BMI body mass index, *DM* diabetes mellitus, *HTN* hypertension, *HLP* hyperlipidemia, *CKD* chronic kidney disease, *AMI* acute myocardial infarction, *STK* stroke, *CAD* coronary artery disease, *HF* heart failure, *Afib* atrial fibrillation, *COPD* chronic obstructive pulmonary disease

under the curve (AUC), sensitivity (Sens.), specificity (Spec.), positive predictive value (PPV), and negative predictive value (NPV) were also calculated. The operating point was selected based on the maximum of Youden's index in the training set. Finally, we used multivariable Cox proportional hazards models to analyze the relationship between the baseline characteristics and the outcomes of interest. Hazard ratios (HRs) and 95% confidence intervals (95% CIs) were used for comparisons. Statistical analysis was carried out using the software environment R version 3.4.4. We used a significance level of $p < 0.05$ throughout the analysis.

Results

Baseline characteristics

Patient characteristics of the development, tuning, internal validation and external validation cohorts are shown in Table 1. Patients in the development cohort were more likely to be female and had a lower BMI and more comorbidities than patients in the validation cohorts. In the development set, 5508 patients (21.5%) had osteoporosis, 7476 patients (29.2%) had osteopenia, and 12,590 patients (49.2%) had normal BMD. In the internal validation cohort, 1142 patients (15.5%) had osteoporosis, 1787 patients (24.3%) had osteopenia, and 4420 patients (60.1%) had normal BMD, while in the external validation cohort, 1071 patients (19.9%) had osteoporosis, 1753 patients (32.6%) had osteopenia, and 2547 patients (47.4%) had normal BMD. In the normal, osteopenia, and osteoporosis groups, it can be observed in Table 2 that as the BMD decreased, the proportion of females and their age increased, while their height and weight decreased; additionally, the proportion of comorbidities increased. Supplementary Fig. 1 shows that the proportion of osteopenia and osteoporosis in different gender and age stratifications was the lowest for osteoporosis in males as age increased, while in females, the proportion of osteoporosis increased with age, with almost 50% of women over 79 years old having symptoms of osteoporosis. To understand the differences in osteoporosis between genders. The gender-stratified results were presented in Supplementary Tables 1 and 2.

Performance of CXR-OP to identify osteoporosis

The algorithm provided discrimination between the osteoporosis and no osteoporosis groups, with an AUC of 0.930 and corresponding sensitivity of 92.9%, specificity of 78.8%, positive predictive value of 44.6%, negative predictive value of 98.4% and F score of 0.603 in the

internal validation set and an AUC of 0.892 and corresponding sensitivity of 92.9%, specificity of 68.1%, positive predictive value of 42.1%, negative predictive value of 97.5% and F score of 0.579 in the external validation set, as shown in Fig. 2. In addition to the performance of the 10-crop approach, the results for the 1-crop and different BMD parts are presented in Supplementary Figs. 2 and 3. Notably, in the spine region, the AUC improved further to 0.940 in the internal validation set.

The model performance was further stratified by hospital department, age group and sex. Looking at different hospital departments, the health check center had the highest AUCs of 0.948 and 0.950 in the internal and external validation sets, respectively. In terms of gender stratification, the AUC was better for males than females. Among age stratifications, the AUC was highest for those under 60 years old. Finally, among the combinations of sex and age, males under 60 years old had the highest AUCs of 0.942 and 0.933 in the internal and external validation sets, respectively, as shown in Fig. 3. Furthermore, to provide

Table 2 Baseline characteristics in different stages of osteoporosis

	Normal	Osteopenia	Osteoporosis	p value
GENDER				<0.001
Female	8526(49.8%)	6043(78.0%)	4207(88.8%)	
Male	8579(50.2%)	1705(22.0%)	531(11.2%)	
Age (years)	51.86 ± 14.40	63.92 ± 13.08	72.42 ± 11.50	<0.001
Height (cm)	163.40 ± 8.57	158.17 ± 7.71	154.70 ± 6.86	<0.001
Weight (kg)	66.36 ± 12.81	59.34 ± 10.44	53.78 ± 9.78	<0.001
BMI (kg/m ²)	24.58 ± 3.88	23.59 ± 3.71	22.52 ± 3.80	<0.001
BMD (T score)	0.56 ± 1.13	-1.72 ± 0.40	-3.15 ± 0.56	<0.001
Disease history				
AMI	58(0.3%)	46(0.6%)	33(0.7%)	0.001
STK	689(4.0%)	763(9.8%)	613(12.9%)	<0.001
CAD	1430(8.4%)	1167(15.1%)	801(16.9%)	<0.001
Afib	173(1.0%)	168(2.2%)	175(3.7%)	<0.001
HF	315(1.8%)	327(4.2%)	306(6.5%)	<0.001
DM	1662(9.7%)	1353(17.5%)	864(18.2%)	<0.001
HTN	339(2.0%)	253(3.3%)	216(4.6%)	<0.001
CKD	561(3.3%)	583(7.5%)	528(11.1%)	<0.001
HLP	3253(19.0%)	2439(31.5%)	1407(29.7%)	<0.001
COPD	1303(7.6%)	1149(14.8%)	910(19.2%)	<0.001

BMI body mass index, *DM* diabetes mellitus, *HTN* hypertension, *HLP* hyperlipidemia, *CKD* chronic kidney disease, *AMI* acute myocardial infarction, *STK* stroke, *CAD* coronary artery disease, *HF* heart failure, *Afib* atrial fibrillation, *COPD* chronic obstructive pulmonary disease

a clearer insight into the regions observed by the AI during predictions, we also generated visual heatmaps to illustrate the significant image information in Supplementary Fig. 4.

Prediction of long-term risk of mortality

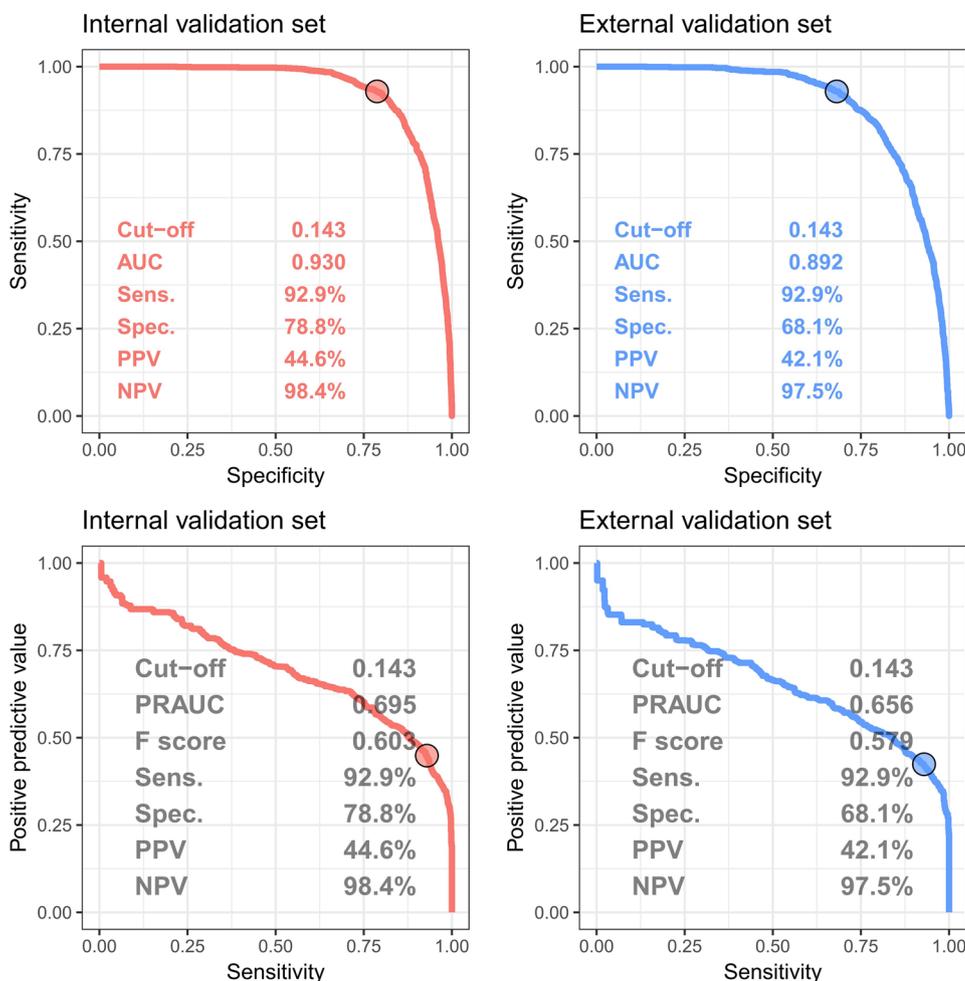
Figure 4 shows the prognostic value in patients stratified by DXA or CXR-Osteoporosis to emphasize the additional prognostic value of CXR-OP. In the traditional DXA examination of bone density, the incidence of all-cause mortality was 7.0% at 2 years and 19.6% at 8 years in the osteoporosis group in the internal validation set, which was not significantly higher than that in the no osteoporosis group (1.3% and 4.6%), with an adjusted HR of 1.39 (95% CI: 0.96–2.01). There was no significant difference observed in the external validation set. We further analyzed the CXR-OP classification and found that the incidence of all-cause mortality was 6.0% at 2 years and 17.3% at 8 years in the osteoporosis group in the internal validation set, which was significantly higher than that in the no osteoporosis group (0.8% and 3.3%),

Fig. 3 The AUCs of ROC curves of DLM predictions based on CXR to detect osteoporosis in different sex and age stratifications. ER: Emergency Room; IPD: Inpatient Department; OPD: Outpatient Department; PEC: Physical Examination Center

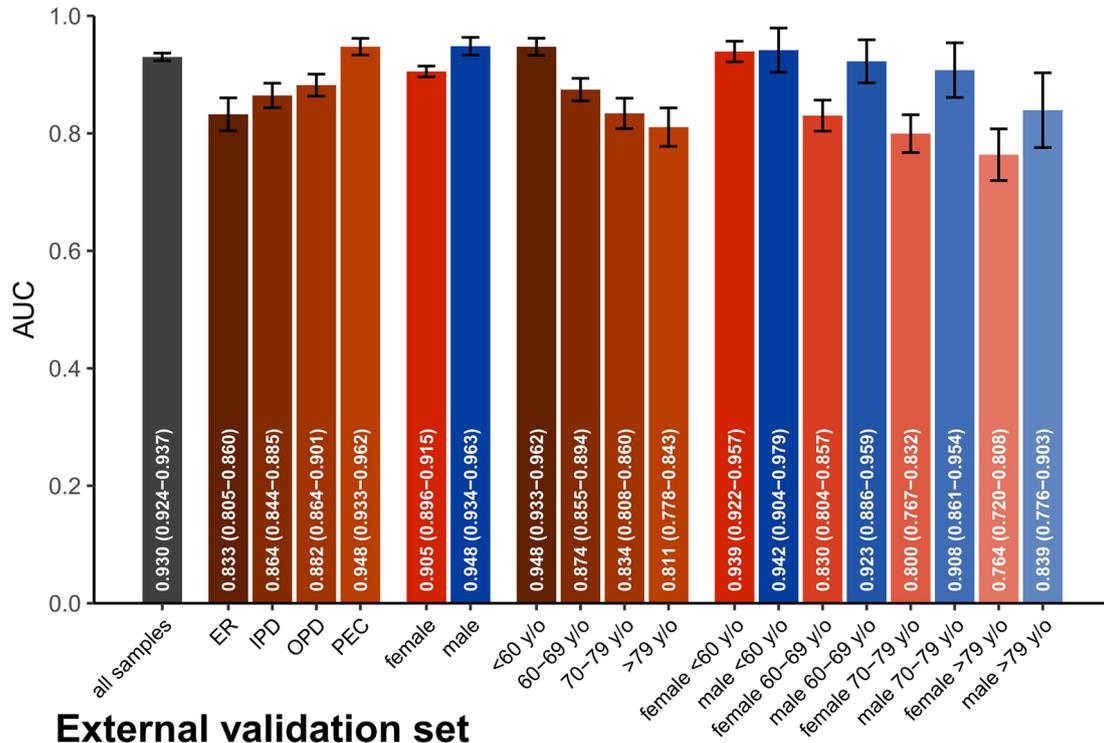
with an adjusted HR of 2.59 (95% CI: 1.83–3.67). This relationship was also validated in the external validation set. The demographic based on mortality status in the internal and external validation sets were presented in Supplementary Table 3.

Figure 5 shows the risk matrices of different CXR-OP and DXA on all-cause mortality. Stratified by DXA, osteoporosis predicted no higher all-cause mortality (HRi 1.08, 95% CI 0.83–1.42 [P=0.561]; HRe 0.93, 95% CI 0.72–1.19 [P=0.559]) than no osteoporosis in either validation cohort. Importantly, patients with CXR-OP(+) significantly contributed to higher risks of all-cause mortality (HRi 2.53, 95% CI 1.76–3.62; HRe 1.85, 95% CI 1.37–2.48) compared with those with CXR-OP(-). CXR-OP independently provided the ability of risk stratification on adverse outcomes.

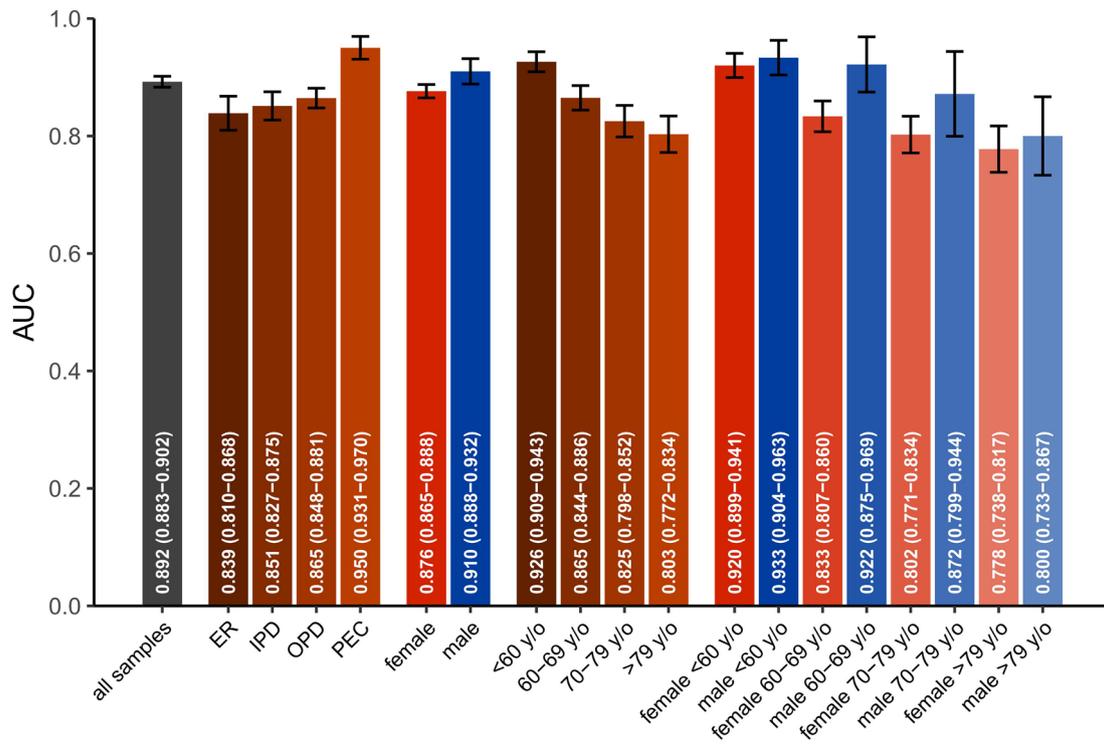
Fig. 2 ROC curves of DLM predictions based on CXR to detect osteoporosis. Osteoporosis is defined as an actual T score of ≤ -2.5 . The operating point was selected based on the maximum of Youden's index in the tuning set and presented using a circle mark, and the area under the ROC curve (AUC), sensitivity (Sens.), specificity (Spec.), positive predictive value (PPV), and negative predictive value (NPV) were calculated



Internal validation set

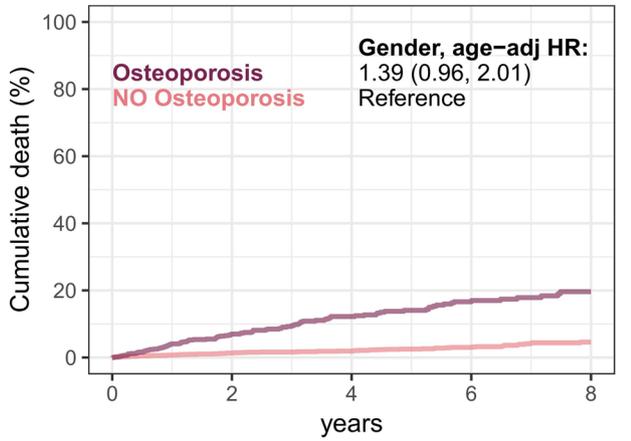


External validation set



Internal validation set

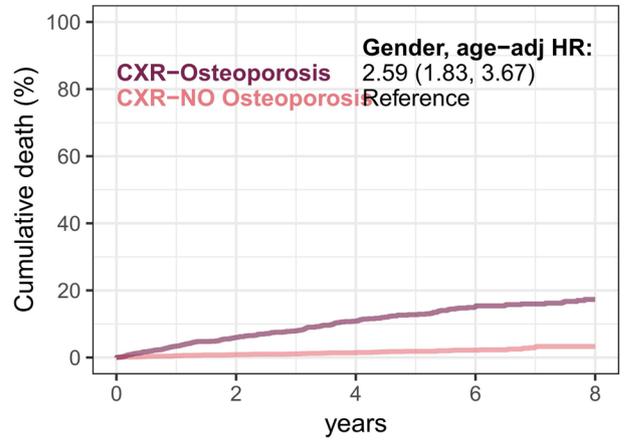
All-cause mortality



Number at risk/event rate (%)

1128 (0.0%)	603 (7.0%)	360 (12.2%)	227 (17.0%)	100 (19.6%)
4388 (0.0%)	2169 (1.3%)	1439 (1.9%)	856 (3.1%)	419 (4.6%)

All-cause mortality

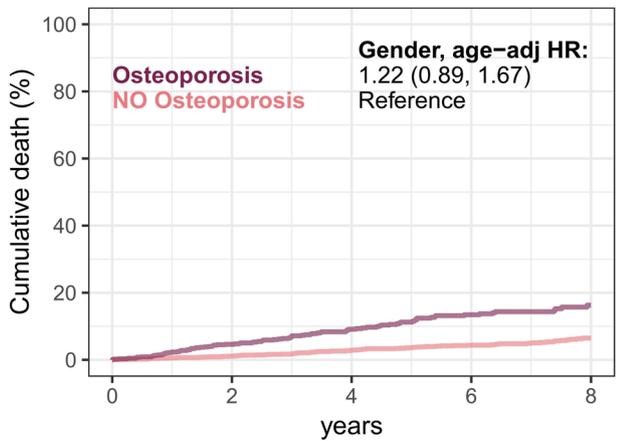


Number at risk/event rate (%)

2344 (0.0%)	1319 (6.0%)	830 (10.8%)	521 (15.4%)	255 (17.3%)
4931 (0.0%)	2452 (0.8%)	1633 (1.4%)	1009 (2.2%)	497 (3.3%)

External validation set

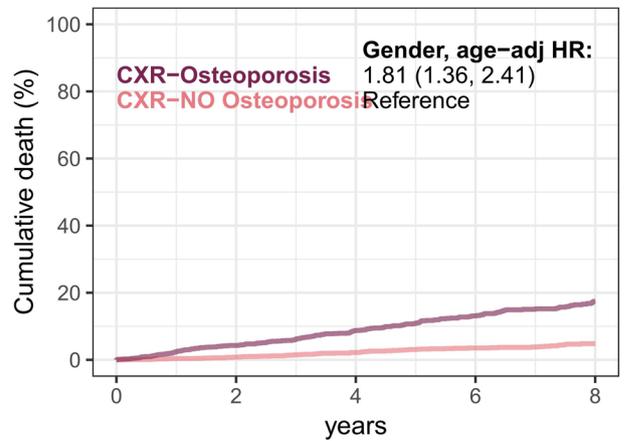
All-cause mortality



Number at risk/event rate (%)

1071 (0.1%)	709 (4.6%)	467 (9.1%)	309 (13.4%)	139 (16.3%)
2547 (0.0%)	1932 (1.1%)	1397 (2.9%)	944 (4.4%)	518 (6.5%)

All-cause mortality



Number at risk/event rate (%)

2365 (0.0%)	1691 (4.3%)	1158 (8.8%)	769 (13.1%)	400 (17.4%)
3006 (0.0%)	2277 (0.8%)	1658 (2.2%)	1135 (3.5%)	623 (4.8%)

Fig. 4 Long-term incidence of developing mortality events stratified by DXA or CXR-Osteoporosis. The analyses were conducted in both the internal and external validation sets. The tables show the at-risk

populations and cumulative risk for the given time intervals in each risk stratification

Therefore, we proceeded to validate BMD in patients who had not undergone DXA examination. The prediction of the long-term development of all-cause mortality in patients

stratified by CXR-OP after adjustment for age and sex is shown in Fig. 6. The incidence of all-cause mortality was 7.3% and 17.5% for 2 and 8 years, respectively, in those

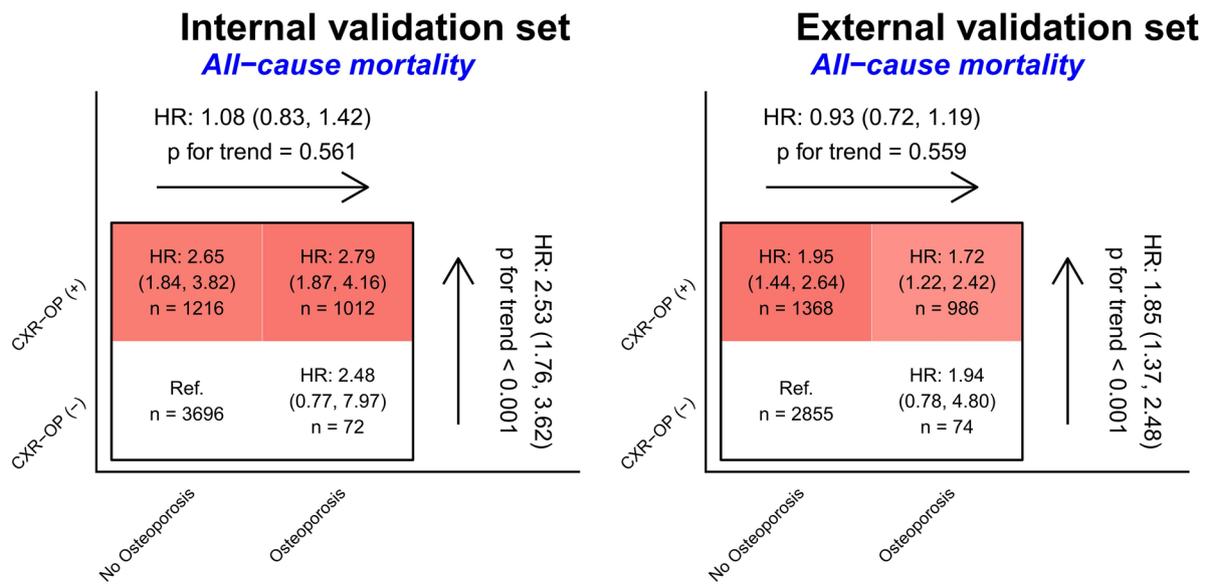


Fig. 5 Risk matrices of long-term all-cause mortality stratified by CXR-OP and DXA. The hazard ratios were based on the Cox proportional hazards model adjusted for sex and age. The color gradient represents the risk of the corresponding group, and nonsignificant

results are shown as white. CXR-OP, deep-learning model to identify osteoporosis via chest X-ray; HR, hazard ratio (with 95% confidence limits)

stratified as CXR-Osteoporosis, compared with 2.0% and 5.2% for 2 and 8 years in those stratified as CXR-No Osteoporosis with an adjusted HR of 1.67 (95% CI: 1.61–1.72) in

patients without a DXA examination. Moreover, to understand the demographic differences between the CXR-Osteoporosis and CXR-No Osteoporosis groups, we presented the demographic comparison in Supplementary Table 4.

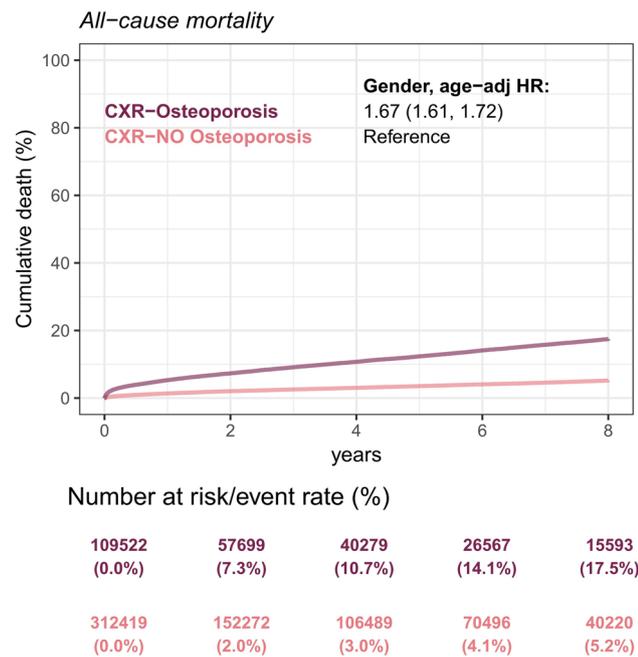


Fig. 6 Long-term incidence of developing mortality events stratified by CXR-Osteoporosis in patients without a DXA examination. The analyses were conducted in both the internal and external validation sets. The tables show the at-risk populations and cumulative risk for the given time intervals in each risk stratification

Discussion

This study developed a DLM AI model based on CXR that can predict BMD T scores and classify high-risk osteoporosis patients. We also extended the application by identifying potential high-risk osteoporosis patients to warn them of their future high risk of death. For patients who have not undergone DXA examination for bone density, their risk of osteoporosis can be detected early through the existing CXR and AI model. The AI-enabled prediction of BMD T scores can be integrated into health information systems as a potential osteoporosis risk assessment, without additional manpower, to enhance risk awareness to health care providers.

In another study that used DXA as the reference for bone density, opportunistic osteoporosis screening tools based on CT attenuation of the spine were used, which achieved an AUC of 0.83 [18]. A machine learning-based T score simulation also showed an accuracy of 0.82 [19], while using pelvis/lumbar spine radiographs for bone density assessment yielded AUCs ranging from 0.92 to 0.97 [29]. In comparison, our tool exhibited a strong correlation with the gold standard DXA-measured BMD in both internal and external testing sets and demonstrated excellent

discriminatory performance in classifying osteoporosis (with AUCs of 0.93 and 0.89, respectively).

In terms of performance in different regions, especially in the spine, the AUC can be enhanced further, reaching 0.940. In women aged 50 to 60 years with more than 3 years of menopause, we tend to observe lower T-scores in the spine as opposed to the hip [35]. Differences in bone loss across various parts of the body and the ratio of cortical to cancellous bones contribute to variations in BMD. Cancellous bone, with its higher turnover rate, tends to be lost earlier than cortical bone. This suggests that the presence of more cancellous bone in the spine may explain the earlier loss of bone matrix in early-stage osteopenia and a more pronounced discrepancy in late-stage osteoporosis [36]. Typically, screening for osteoporosis in chest radiographs relies on assessing the clavicles, ribs, and spine [37–39]. Consequently, Fig. S4 reveals that our model predominantly concentrates on distinguishing the rib area, potentially causing misinterpretations as lung fields and the CV system. Earlier research has emphasized a significant correlation between rib density and overall bone density ($r=0.86$) [40], as well as correlations ranging from 0.67 to 0.75 among lumbar spine, femoral neck, and whole-body bone density [41]. Furthermore, our AI model was particularly precise in predicting osteoporosis in males and individuals below 60 years of age. In contrast, similar studies to ours had a lower AUC of 0.84 in predicting the risk of osteoporosis [42].

In addition to predicting osteoporosis risk, our model could also predict the risk of osteopenia with an AUC of approximately 0.85 (data not shown). This is because among the participants who underwent bone densitometry, those diagnosed with osteoporosis had a higher fracture rate, but there was a larger number of patients diagnosed with osteopenia. Consequently, the total number of fractures was higher in the group diagnosed with osteopenia than in the group diagnosed with osteoporosis [43]. Medical guidelines recommend further examination or therapeutic interventions for osteopenia or osteoporosis [11, 44, 45].

As individuals age and their bodies undergo the aging process, they become more susceptible to chronic diseases [46], resulting in a lower quality of life [47] and a higher risk of osteoporosis [48] and fractures [49]. Therefore, when treating older adults with multimorbidity, it is important to consider the competing risk of death when assessing the risks and benefits of treatment [50]. Clinicians may use validated prediction models, such as the FRAX tool [51] and Garvan fracture risk calculator [52], to compare the absolute risk of fracture with the risk of death. By doing so, they can make more informed decisions regarding treatment options for their patients. In our study, we used medical record data to assess the risk of mortality in patients with osteoporosis. When classified according to the reference standard of DXA, patients with osteoporosis had a risk of mortality of

1.39 (95% CI: 0.96–2.01) in both the internal and external validation sets. However, when classified using the CXR-OP system, the AI model predicted a higher risk of mortality in individuals with a higher risk of osteoporosis, with an HR of 2.59 (95% CI: 1.83–3.67), which was also validated in the external validation set. In Fig. 6, analogous outcomes were found in the cohort that did not undergo DXA examinations.

Supplementary Table 4 illustrates that within the CXR-OP group, there is a higher representation of females, older individuals, lower stature, reduced body weight, lower BMI, and a greater prevalence of various medical histories. Despite the CXR-OP group in Supplementary Table 4 appearing roughly 20 years older than the CXR-NO OP group, we conducted a gender and age-stratified analysis on the internal/external validation sets depicted in Fig. 3 to validate the model's predictive capabilities. Figure 3 reveals optimal predictive performance in the population under 60 years old (AUC: 0.948/0.926), with a gradual decline in performance with increasing age. Moreover, in Figs. 4 to 6 our Cox proportional hazards model was adjusted for both sex and age. Consequently, the model does not forecast age but rather showcases discriminative abilities for individuals with osteoporosis. As a result, our model not only predicts the risk of osteoporosis but also assesses the risk of mortality.

This retrospective study has several limitations. Firstly, while CXRs were gathered from a medical center, validating CXR-OP's accuracy and applicability in the community necessitates prospective studies. Secondly, despite attempts to address class imbalance and overfitting, careful evaluation is essential for the DLM's generalization, with further studies required for confirmation. Thirdly, the DLM-identified CXR traits remain unspecified due to automated feature creation. Lastly, CXR-OP's osteoporosis detection uses a limited T score range (≤ -2.5 and > -2.5), differing from the World Health Organization's DXA-based criteria [53]. Osteopenia, less severe but clinically relevant, warrants regular bone density monitoring to prevent progression. Despite its CXR-based osteoporosis screening role, CXR-OP doesn't consider actual density or fracture risk in a patient's clinical context.

Conclusion

We developed a DLM from a large set of CXRs validated by DXA to identify osteoporosis. This novel strategy provides a common, feasible, and affordable method to assist physicians in the early identification of individuals with a risk of osteoporosis. AI-enabled CXR may permit the addition of significant prognostic implications for osteoporosis and serve as a screening tool to improve quality of life and reduce the risk of death.

Abbreviations AI: Artificial Intelligence; AUC: Area Under the Curve; AMI: Acute Myocardial Infarction; Afib: Atrial fibrillation; BMD: Bone Mineral Density; CT: Computed Tomography; CXR: Chest X-Ray; CVD : Cardiovascular Disease; CV: Cardiovascular; CAD: Coronary Artery Disease; DXA: Dual energy X-ray Absorptiometry; DLM: Deep Learning Model; DICOM: Digital Imaging and Communications in Medicine; HR: Hazard Ratio; HF: Heart failure; ICD: International Classification of Diseases; NPV: Negative Predictive Value; OP: Osteoporosis; PPV: Positive Predictive Value; ROC: Receiver Operating Characteristic; STK: Stroke; Sens.: Sensitivity; Spec.: Specificity; WHO: World Health Organization

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s10916-023-02030-2>.

Author contributions All authors participated in designing the study, generating hypotheses, interpreting the data, and critically reviewing the paper. DJT and WHF wrote the first draft, and CL, CSL, CCL, and CHW contributed substantially to writing subsequent versions. DJT designed and conducted statistical analyses with support from CL. All authors had full access to all the data in the study and accepted responsibility for the decision to submit for publication. DJT and WHF verified all the data used in this study. The corresponding author (WHF) attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Funding National Science and Technology Council, NSTC 112-2222-E-030 -002 -MY2, NSTC 112-2321-B-016-003, Ministry of Science and Technology, Taiwan, MOST110-2314-B-016-010-MY3.

Data availability The data analyzed in this study is not publicly available due to privacy and security concerns. The data may be shared with a third party upon execution of data sharing agreement for reasonable requests, such requests should be addressed to WHF (e-mail: rumaf.fang@gmail.com) or DJT.

Declarations

Ethics approval The Tri-Service General Hospital, Taipei, Taiwan, conducted the ethical review of this study (IRB No. C202105049). The institutional review board agreed to waive individual consent since the data were collected retrospectively and analyzed on the intranet.

Competing interests The authors declare no competing interests.

References

- Gharib, H., et al., *Associazione Medici Endocrinologi, and European Thyroid Association medical guidelines for clinical practice for the diagnosis and management of thyroid nodules*. *Endocr Pract*, 2010. **16**(Suppl 1): p. 1-43.
- Organization, W.H. *WHO scientific group on the assessment of osteoporosis at primary health care level*. in *Summary meeting report*. 2004.
- Curry, S.J., et al., *Screening for osteoporosis to prevent fractures: US Preventive Services Task Force recommendation statement*. *Jama*, 2018. **319**(24): p. 2521-2531.
- Shao, C.-J., et al., *A nationwide seven-year trend of hip fractures in the elderly population of Taiwan*. *Bone*, 2009. **44**(1): p. 125-129.
- Wooltorton, E., *Osteoporosis treatment: raloxifene (Evista) and stroke mortality*. *Cmaj*, 2006. **175**(2): p. 147.
- Bliuc, D., et al., *Mortality risk associated with low-trauma osteoporotic fracture and subsequent fracture in men and women*. *Jama*, 2009. **301**(5): p. 513-21.
- Ganry, O., et al., *Bone mass density and risk of breast cancer and survival in older women*. *Eur J Epidemiol*, 2004. **19**(8): p. 785-92.
- Cai, S., et al., *Bone mineral density and osteoporosis in relation to all-cause and cause-specific mortality in NHANES: A population-based cohort study*. *Bone*, 2020. **141**: p. 115597.
- Yamamoto, N., et al., *Deep learning for osteoporosis classification using hip radiographs and patient clinical covariates*. *Bio-molecules*, 2020. **10**(11): p. 1534.
- Mueller, D. and A. Gandjour, *Cost-effectiveness of using clinical risk factors with and without DXA for osteoporosis screening in postmenopausal women*. *Value in Health*, 2009. **12**(8): p. 1106-1117.
- Orimo, H., et al., *Japanese 2011 guidelines for prevention and treatment of osteoporosis—executive summary*. *Archives of osteoporosis*, 2012. **7**: p. 3-20.
- Sedlak, C.A., M.O. Doheny, and S.L. Jones, *Osteoporosis education programs: changing knowledge and behaviors*. *Public health nursing*, 2000. **17**(5): p. 398-402.
- Sato, M., et al., *Bone fractures and feeling at risk for osteoporosis among women in Japan: patient characteristics and outcomes in the National Health and Wellness Survey*. *Archives of Osteoporosis*, 2014. **9**: p. 1-9.
- Compston, J.E., M.R. McClung, and W.D. Leslie, *Osteoporosis*. *Lancet*, 2019. **393**(10169): p. 364-376.
- Curtis, J.R., et al., *Longitudinal trends in use of bone mass measurement among older americans, 1999-2005*. *J Bone Miner Res*, 2008. **23**(7): p. 1061-7.
- Davis, S.R., et al., *Simplifying screening for osteoporosis in Australian primary care: the Prospective Screening for Osteoporosis: Australian Primary Care Evaluation of Clinical Tests (PROSPECT) study*. *Menopause*, 2011. **18**(1): p. 53-9.
- Yasaka, K., et al., *Prediction of bone mineral density from computed tomography: application of deep learning with a convolutional neural network*. *European radiology*, 2020. **30**: p. 3549-3557.
- Pickhardt, P.J., et al., *Opportunistic screening for osteoporosis using abdominal computed tomography scans obtained for other indications*. *Annals of internal medicine*, 2013. **158**(8): p. 588-595.
- Krishnaraj, A., et al., *Simulating dual-energy X-ray absorptiometry in CT using deep-learning segmentation cascade*. *Journal of the American College of Radiology*, 2019. **16**(10): p. 1473-1479.
- Dagan, N., et al., *Automated opportunistic osteoporotic fracture risk assessment using computed tomography scans to aid in FRAX underutilization*. *Nature medicine*, 2020. **26**(1): p. 77-82.
- Benhamou, C.-L., et al., *Fractal analysis of radiographic trabecular bone texture and bone mineral density: two complementary parameters related to osteoporotic fractures*. *Journal of bone and mineral research*, 2001. **16**(4): p. 697-704.
- LeCun, Y., Y. Bengio, and G. Hinton, *Deep learning*. *nature*, 2015. **521**(7553): p. 436-444.
- Lindsey, R., et al., *Deep neural network improves fracture detection by clinicians*. *Proceedings of the National Academy of Sciences*, 2018. **115**(45): p. 11591-11596.
- He, K., et al. *Delving deep into rectifiers: Surpassing human-level performance on imagenet classification*. in *Proceedings of the IEEE international conference on computer vision*. 2015.
- Gulshan, V., et al., *Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs*. *Jama*, 2016. **316**(22): p. 2402-2410.
- Ardila, D., et al., *End-to-end lung cancer screening with three-dimensional deep learning on low-dose chest computed tomography*. *Nature medicine*, 2019. **25**(6): p. 954-961.

27. Smets, J., et al., *Machine learning solutions for osteoporosis—a review*. Journal of bone and mineral research, 2021. **36**(5): p. 833-851.
28. Nguyen, T.P., et al., *A novel approach for evaluating bone mineral density of hips based on Sobel gradient-based map of radiographs utilizing convolutional neural network*. Computers in Biology and Medicine, 2021. **132**: p. 104298.
29. Hsieh, C.-I., et al., *Automated bone mineral density prediction and fracture risk assessment using plain radiographs via deep learning*. Nature communications, 2021. **12**(1): p. 5472.
30. Zhang, B., et al., *Deep learning of lumbar spine X-ray for osteopenia and osteoporosis screening: A multicenter retrospective cohort study*. Bone, 2020. **140**: p. 115561.
31. Jang, M., et al., *Opportunistic osteoporosis screening using chest radiographs with deep learning: Development and external validation with a cohort dataset*. Journal of Bone and Mineral Research, 2022. **37**(2): p. 369-377.
32. Sato, Y., et al., *Deep Learning for Bone Mineral Density and T-Score Prediction from Chest X-rays: A Multicenter Study*. Bio-medicines, 2022. **10**(9).
33. Chang, C.H., et al., *Electrocardiogram-based heart age estimation by a deep learning model provides more information on the incidence of cardiovascular disorders*. Front Cardiovasc Med, 2022. **9**: p. 754909.
34. Liu, W.T., et al., *A deep-learning algorithm-enhanced system integrating electrocardiograms and chest X-rays for diagnosing aortic dissection*. Can J Cardiol, 2022. **38**(2): p. 160-168.
35. Seok, H., et al., *High prevalence of spine–femur bone mineral density discordance and comparison of vertebral fracture risk assessment using femoral neck and lumbar spine bone density in Korean patients*. Journal of bone and mineral metabolism, 2014. **32**: p. 405-410.
36. El Maghraoui, A., et al., *Prevalence and risk factors of discordance in diagnosis of osteoporosis using spine and hip bone densitometry*. Annals of the rheumatic diseases, 2007. **66**(2): p. 271-272.
37. Kumar, D.A. and M. Anburajan, *The role of hip and chest radiographs in osteoporotic evaluation among south Indian women population: a comparative scenario with DXA*. Journal of endocrinological investigation, 2014. **37**: p. 429-440.
38. Holcombe, S.A., et al., *Measuring rib cortical bone thickness and cross section from CT*. Med Image Anal, 2018. **49**: p. 27-34.
39. Chen, H., et al., *Age-related changes in trabecular and cortical bone microstructure*. Int J Endocrinol, 2013. **2013**: p. 213234.
40. Yao, W.J., et al., *Differential Changes in Regional Bone Mineral Density in Healthy Chinese: Age-Related and Sex-Dependent*. Calcified Tissue International, 2001. **68**(6): p. 330-336.
41. Rajaei, A., et al., *Correlating Whole-Body Bone Mineral Densitometry Measurements to Those From Local Anatomical Sites*. Iran J Radiol, 2016. **13**(1): p. e25609.
42. Sato, Y., et al., *Deep learning for bone mineral density and T-score prediction from chest X-rays: A multicenter study*. Bio-medicines, 2022. **10**(9): p. 2323.
43. Siris, E.S., et al., *Bone mineral density thresholds for pharmacological intervention to prevent fractures*. Archives of internal medicine, 2004. **164**(10): p. 1108-1112.
44. Kanis, J.A., et al., *European guidance for the diagnosis and management of osteoporosis in postmenopausal women*. Osteoporosis international, 2013. **24**(1): p. 23-57.
45. Camacho, P.M., et al., *American Association of Clinical Endocrinologists/American College of Endocrinology clinical practice guidelines for the diagnosis and treatment of postmenopausal osteoporosis—2020 update*. Endocrine Practice, 2020. **26**: p. 1-46.
46. Molarius, A. and S. Janson, *Self-rated health, chronic diseases, and symptoms among middle-aged and elderly men and women*. Journal of Clinical Epidemiology, 2002. **55**(4): p. 364-370.
47. Hallberg, I., et al., *Health-related quality of life after osteoporotic fractures*. Osteoporosis International, 2004. **15**: p. 834-841.
48. Hannan, M.T., et al., *Risk factors for longitudinal bone loss in elderly men and women: the Framingham Osteoporosis Study*. Journal of Bone and Mineral Research, 2000. **15**(4): p. 710-720.
49. De Laet, C.E. and H.A. Pols, *Fractures in the elderly: epidemiology and demography*. Best Practice & Research Clinical Endocrinology & Metabolism, 2000. **14**(2): p. 171-179.
50. Berry, S.D., et al., *Competing risk of death: an important consideration in studies of older adults*. J Am Geriatr Soc, 2010. **58**(4): p. 783-7.
51. Kanis, J.A., et al., *FRAX® and its applications to clinical practice*. Bone, 2009. **44**(5): p. 734-743.
52. Nguyen, N.D., et al., *Development of a nomogram for individualizing hip fracture risk in men and women*. Osteoporosis International, 2007. **18**: p. 1109-1117.
53. *Assessment of fracture risk and its application to screening for postmenopausal osteoporosis. Report of a WHO Study Group*. World Health Organ Tech Rep Ser, 1994. **843**: p. 1–129.

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

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.