Artificial Intelligence-enabled Chest X-ray Detects Osteoporosis with Bone Mineral Density and identifies the mortality events

Dung-Jang Tsai
Department of Statistics and Information Science, Fu Jen Catholic University, New Taipei City, Taiwan, R.O.C.

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

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

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

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

Wen-Hui Fang (✉ rumaf.fang@gmail.com)
Department of Family and Community Medicine, Department of Internal Medicine, Tri-Service General Hospital, National Defense Medical Center, Taipei, Taiwan, R.O.C.

Research Article

Keywords: Artificial intelligence, bone mass density, T score, chest X-ray, deep learning

Posted Date: August 28th, 2023

DOI: https://doi.org/10.21203/rs.3.rs-2959194/v2

License: This work is licensed under a Creative Commons Attribution 4.0 International License. Read Full License
Abstract

SUMMARY
A deep learning model was developed to identify osteoporosis from chest X-ray features with high accuracy in internal and external validation. It has significant prognostic implications, identifying individuals at higher risk of all-cause mortality. This AI-enabled chest X-ray strategy may function as an early detection screening tool for osteoporosis.

OBJECTIVE
The aim of this study was to develop a deep learning model (DLM) to identify osteoporosis via chest X-ray features and investigate the performance and clinical implications.

METHOD
This study collected 48,353 CXRs with the corresponding T score according to DXA from the academic medical center. Among these, 35,633 CXRs were used to identify 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.

RESULTS
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.

CONCLUSION
Our DLM uses chest X-rays 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.

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 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 mortality[8].

Currently, the most reliable method of diagnosing osteoporosis is to measure BMD in the hip and lumbar spine using 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 score ≤ −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].

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[14], classify osteoporosis[15], simulate DXA T scores[16], and predict fracture risk[17]. 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[18]. 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[19], which is essential for clinical applications such as fracture detection[20, 21], retinopathy grading[22], and lung nodule identification[23]. Recent advancements in orthopedic research have paved the way for the application of DLMs in osteoporosis screening[24]. Previous studies have shown the feasibility of diagnosing osteoporosis based on radiographs of the lumbar spine and hip joint[25, 26], as well as measuring the BMD (g/cm2) of these sites from radiographs[9, 27]. Furthermore, two studies have utilized chest X-rays for diagnosing osteoporosis[28, 29].
Our hypothesis is that a DLM could accurately predict BMD by analyzing chest radiography. With this in mind, we aimed to develop a DLM that could predict BMD (in g/cm²) and diagnosis based on T scores (normal, osteopenia, and osteoporosis) using chest X-rays, age, and sex. To achieve this, we trained our model using a large dataset gathered from a medical center. Our ultimate goal is to create a model with excellent predictive performance that would allow chest X-rays to be used 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 365 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 chest X-rays for DLM training. A total of 4,878 patients were assigned to the tuning set. These patients provided 10,059 chest X-rays 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 chest X-ray 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/cm²) 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 online 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, Ninth Revision and Tenth Revision (ICD-9 and ICD-10, respectively) described previously.[30] 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 cardiovascular (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. 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 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[31] 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. The performance of the DLMs was evaluated by the receiver operating characteristic (ROC) curve for implant pacemaker analysis, and the area 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. Figure S1 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.

Performance of CXR-OP to identify osteoporosis

The algorithm provided discrimination between the osteoporosis and normal 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 cohort 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 cohort, as shown in Figure 2.

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 Figure 3.

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%), with an adjusted HR of 2.59 (95% CI: 1.83-3.67). This relationship was also validated in the external validation set.

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.

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 Figure 6. The incidence of all-cause mortality was 7.3% and 17.5% for 2 and 8 years, respectively, in those 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.

**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[15]. A machine learning-based T score simulation also showed an accuracy of 0.82[16], while using pelvis/lumbar spine radiographs for bone density assessment yielded AUCs ranging from 0.92 to 0.97[32]. 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). 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[33]. 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[34]. Medical guidelines recommend further examination or therapeutic interventions for osteopenia or osteoporosis[11, 35, 36].

As individuals age and their bodies undergo the aging process, they become more susceptible to chronic diseases[37], resulting in a lower quality of life[38] and a higher risk of osteoporosis[39] and fractures[40]. 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[41]. Clinicians may use validated prediction models, such as the FRAX tool[42] and Garvan fracture risk calculator[43], 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. Therefore, 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[44]. 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.

Declarations

Conflict of Interest
Dung-Jang Tsai, Chin Lin, Chin-Sheng Lin, Chia-Cheng Lee, Chih-Hung Wang and Wen-Hui Fang declare that they have no conflict of interest.

**Funding**

This study was supported by funding from the Ministry of Science and Technology, Taiwan (MOST110-2314-B-016-010-MY3 to C. Lin, NSTC 112-2321-B-016-003 to C.H. Wang and NSTC 112-2222-E-030-002-MY2 to D.J. Tsai)

**References**


**Tables**

Table 1. | Baseline characteristics

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

Table 2. | Baseline characteristics in different stages of osteoporosis

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Osteopenia</th>
<th>Osteoporosis</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
<td></td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Female</td>
<td>8526(49.8%)</td>
<td>6043(78.0%)</td>
<td>4207(88.8%)</td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>8579(50.2%)</td>
<td>1705(22.0%)</td>
<td>531(11.2%)</td>
<td></td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td>51.86±14.40</td>
<td>63.92±13.08</td>
<td>72.42±11.50</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Height</strong></td>
<td>163.40±8.57</td>
<td>158.17±7.71</td>
<td>154.70±6.86</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>Weight</strong></td>
<td>66.36±12.81</td>
<td>59.34±10.44</td>
<td>53.78±9.78</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>BMI</strong></td>
<td>24.58±3.88</td>
<td>23.59±3.71</td>
<td>22.52±3.80</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td><strong>BMD (T score)</strong></td>
<td>0.56±1.13</td>
<td>-1.72±0.40</td>
<td>-3.15±0.56</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

**Disease history**

<table>
<thead>
<tr>
<th></th>
<th>Normal</th>
<th>Osteopenia</th>
<th>Osteoporosis</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>AMI</td>
<td>58(0.3%)</td>
<td>46(0.6%)</td>
<td>33(0.7%)</td>
<td>0.001</td>
</tr>
<tr>
<td>STK</td>
<td>689(4.0%)</td>
<td>763(9.8%)</td>
<td>613(12.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CAD</td>
<td>1430(8.4%)</td>
<td>1167(15.1%)</td>
<td>801(16.9%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Afib</td>
<td>173(1.0%)</td>
<td>168(2.2%)</td>
<td>175(3.7%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HF</td>
<td>315(1.8%)</td>
<td>327(4.2%)</td>
<td>306(6.5%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>DM</td>
<td>1662(9.7%)</td>
<td>1353(17.5%)</td>
<td>864(18.2%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HTN</td>
<td>339(2.0%)</td>
<td>253(3.3%)</td>
<td>216(4.6%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>CKD</td>
<td>561(3.3%)</td>
<td>583(7.5%)</td>
<td>528(11.1%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>HLP</td>
<td>3253(19.0%)</td>
<td>2439(31.5%)</td>
<td>1407(29.7%)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>COPD</td>
<td>1303(7.6%)</td>
<td>1149(14.8%)</td>
<td>910(19.2%)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>


**Figures**
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.
Figure 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.
Figure 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.
Figure 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.
Figure 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).
Figure 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.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1.png