Supplementary methods

DXA measurement

The densitometer is operated by multiple radiology technicians. In CGMH, all technicians need to be certified to operate the densitometer. The scans were analyzed following recommendations issued by the Taiwan Radiological Society⁴⁶ (amended from International Society for Clinical Densitometry, ISCD).⁴⁷ The recommendation briefly: 1. All DXA technicians need to complete the ISCD training course (basic and advanced) and operate at least 120 cases of DXA examination before the independent operation of DXA. 2. All technicians need to obtain the SOP provided by the manufacturer and conduct BMD measurements accordingly. 3. All DXA scanners need to have a detailed SOP at the examination site, which needs to be updated regularly and reviewed by relevant professionals. 4. All DXA scanners need to comply with the local radiation safety guideline. 5. Spine phantom BMD measurement is performed regularly to document the stability of DXA performance over time. BMD values must be maintained within an error of ±1.5%. A monitoring plan is needed for a correction approach when the error has been exceeded. 6. All DXA technicians need to establish a personal least significant change (LSC). The current recommendation is that the personal LSC should be within 5.3% for the lumbar spine, 5.0% for the whole hip, and 6.9% for hip neck BMD measurement.

Deep Adaptive Graph (DAG)

In DAG, the anatomical landmarks are formulated as a graph, $G = (V, E, F)$, where the vertices V represent the landmarks, the edges E represent the relationships between them, and the features F encode visual patterns in the neighborhoods of the vertices. For a specific input image, the graph vertices are first initialized in the image using the mean shape of the anatomy, and a neural network is used to

displace the vertices from the initial position to the target anatomy in the image. By formulating the anatomical landmarks as a graph and modeling their displacements by the convolutional neural network– graph convolutional network (GCN), DAG can effectively exploit the structural information and shape prior to the anatomical landmarks. Therefore, DAG provides robust and accurate anatomical landmark detection on both hip and spine radiographs.

The neural network consists of a convolutional neural network to encode the input image to produce graph features F, and a GCN to process the graph to locate its vertices. Specifically, the GCN consists of two parts: a global transformation GCN and multiple local refinement GCNs. The global transformation GCN produces an affine transformation matrix, M, which brings the initial graph vertices closer to the target. The transformed vertices are written as follows:

Equation (4): $V^1 = \{v_i^1\} = \{Mv_i^0\}$,

where $V^0 = {v_i}^0$ and $V^1 = {v_i}^1$ denote the initial graph vertices before and after the estimated affine transformation, respectively. The local refinement GCNs then iteratively estimate the displacements of the graph vertices $\,V^1.$ In each iteration, the vertices are displaced as follows:

Equation (5):
$$
v_i^{t+1} = v_i^t + \Delta v_i^t
$$
,

where Δv_i^t is the displacement estimated by the local refinement GCN at the t-th step.

The training loss is calculated for both the global transformation GCN and the local refinement GCNs.

Because the goal of global transformation GCN is to locate the anatomy coarsely, the following margin loss

is used:

Equation (6): L_{global} =
$$
\left[\frac{1}{N} \sum_{i \in N} |v_i^1 - v_i| - m \right]_+
$$

where $[u]_+ = \max(0, u)$; v_i^1 and v_i Denote the globally transformed and ground truth vertices, respectively; and m is a hyperparameter representing a margin that aims to achieve high robustness for coarse landmark detection and forgive small errors. To encourage the local refinement GCNs to learn a precise localization, the L1 loss is directly applied to all vertices after the refinements, written as follows:

Equation (7): L_{local} =
$$
\frac{1}{N} \sum_{i \in N} |v_i^T - v_i|
$$
,

where v_i^T denotes the vertices after the last local refinement GCN. The graph edge weights are treated as learnable parameters, which are initialized randomly at the beginning of training and updated via backpropagation during training.

Implementation details

Deep learning models were developed on a workstation with a single Intel Xeon E5-2650 v4 CPU @ 2.2 GHz, 128 GB RAM, and 4 NVIDIA TITAN V GPUs running Ubuntu 18.04 LTS. All code used in this study was developed in Python v3.6, and deep learning models were implemented using PyTorch v1.3. Image preprocessing was performed using the Python Imaging Library. ImageNet pre-trained weights were used to initialize the backbone network VGG-16 block. The Adam optimizer was used to train the model for 200 epochs with a batch size of 8, a starting learning rate of 1e−4, and a weight decay of 1e−4. The learning rate was reduced to 1e−5 after the first 100 training epochs. The following augmentations are performed during training: 1) color jittering on both the brightness (+-0.2) and contrast (+-0.2). 2) random up-down and leftright flipping, 3) random affine transformation (rotation +-30, shear +- 0.2, translation +- 25 pixels, scaling

+-10%). The trained model was evaluated on the validation set after each training epoch, and the model

with the highest validation correlation coefficient r-value was selected as the best model.

Supplementary Figure 1. Flowchart of the study population.

Supplementary Figure 2 Comparison of the receiver-operator curve between FRAX-AI (Navy blue line) and FRAX-NB tools (Maroon line) to classify high 10-year risk groups for (a) major fracture (>=20%) and (b) hip fracture (>=3%)

(a)

Supplementary Figure 3 Comparison of the precision-recall curve between FRAX-AI (Navy blue line) and FRAX-NB tools (Maroon line) to classify high 10-year risk groups for (a) major fracture (>=20%) and (b) hip

(b)

Supplementary Figure 4 The calibration plots for predicted-measured BMD. (a) 2060 pairs of predictedmeasured hip BMD (2060 patients) and (b) 11027 pairs of predicted-measured lumbar vertebral BMD (3346 patients). The DXA BMD was based on the measurement by the GE Lunar iDXA. Each point represents a data pair of predicted and measure BMD. The points close to the diagonal line suggests good calibration.

(a)

DXA-measured hip BMD (g/cm²)

Supplementary Figure 5 Data flow chart for the real-world clinical test. (a) hip BMD predictions and (b) spine BMD prediction. In this real-world test, we collected all patients receiving plain film for the hip and lateral plain film of the lumbar spine during January–May 2021. These images were fed into our tool and estimate the number (%) of images that can pass quality check, the distribution of predicted BMD, and the categorization based on the threshold reported in table 4.

(a)

Supplementary tables

Supplementary Table 1 Network performance across age and sex subsets

* The number of patients with a 10-year risk of major osteoporotic fracture >=20% is less than 3.

Supplementary Table 2. Patient characteristics of the GE dataset for testing*

* Converted to the Hologic equivalent

** Calculated based on vertebrae with the lowest BMD.

Supplementary Table 3 BMD value conversion equations for GE Lunar to Hologic DXA scanners.

Supplementary Table 4. Summary of performance metrics of the predictive model for BMD* (GE DXA scanner)

* GE BMD were converted to the Hologic equivalent.

** Calculated per eligible vertebrae.

Supplementary Table 5 Discriminatory performance (%) of the predicted BMD to classify hip/lumbar vertebral osteoporosis and high-risk groups for major osteoporotic or hip fractures (GE DXA scanner).

Supplementary Table 6 Model performance in different x-ray machines and tanges of peak kilovolt (kV) for hip and lumbar radiographs.

* Means were compared using student t test. Two-sided p values were reported.

Supplementary Table 8. Summary of performances using different backbone networks for spine BMD prediction (Hologic DXA).

* Means were compared using student t test. Two-sided p values were reported.

Supplementary Table 9. Pseudo code for hip BMD estimation in python style.

```
Input: hip X-ray image, I
def hip bm(I):""""
   Estimate hip BMD from pelvic X-ray images
   """ ""
  lmks = DagModel(I) # Apply the DAG model
  score fx = FranctureModel(I) # Apply the fracture detection model
  score implant = ImplantModel(I) # Apply the implant detection model
  box 1, box r = roi bounding box(lmks) # Generate bounding boxes
  fx l = roi pooling(score fx, box 1) > 0.5 # Fracture classification
  fx r = roi pooling(score fx, box r) > 0.5
  impl l = roi pooling(score impl, box 1) > 0.5 # Implant classification
  impl r = \text{roi pooling}(\text{score impl, box r}) > 0.5if not fx l and not impl l: # if no fracture and implant
     roi l = roi extraction(I, box 1) # extract ROI of the left hip
     bmd l = BmdModel(roi 1) # Apply the BMD model
    else:
      bmd l = None
  in not fx r and not impl r:
     roi r = roi extraction(I, box r)
      bmd r = BmdModel(roi r)
```
else:

 b md r = None

return bmd_l, bmd_r # return left and right hip BMDs

Supplementary Table 10. Pseudo code for spine BMD estimation in python style.

```
Input: Spine X-ray image, I
def spine bm(I):
   """
    Estimate spine BMD from lateral spine X-ray images
   "" ""
   lmks = DagModel(I) # Apply the DAG model
  score fx = FractureModel(I) # Apply the fracture detection model
   score implant = ImplantModel(I) # Apply the implant detection model
   boxes = roi bounding box(lmks) # Generate bounding boxes as a dictionary
   bmd = \{\} for vert in ['L1', 'L2', 'L3', 'L4']:
      fx = roi pooling(score fx, boxes[vert]) > 0.5 # Fracture classification
      impl = roi pooling(score impl, boxes[vert]) > 0.5 # Implant classification
      ap ratio, mid ratio, height ratio = \backslashsix point morph(boxes[vert]) # Calculate 6-point morphology metrics
       if not fx and not impl and \
         ap ratio > 0.8 and mid raio > 0.6 and height ratio > 0.55: # If there is no fracture or implant and the morphology
is normal
         roi = roi extraction(I, boxes[vert])
         bmd[vert] = BmdModel(roi) else:
          bmd[vert] = None
```
return bmd['L1'], bmd['L2'], bmd['L3'], bmd['L4'] # Return 4 L-spine BMDs

Step 1: Open the project page https://gigantum.com/xraybmd/nc-bmd-cpu

It will take quite some time, but it will load eventually.

√ Download Gigantum Client to work locally, for free

There are no compute limits when you are using Gigantum Client locally. Work as often as you want, for as long as you want, on whatever compute resource you choose. Download app here to get started.

xraybmd - nc-bmd-cpu

■ gigantum.com/client/xraybmd/nc-bmd-cpu/jupyterlab \rightarrow C △

 \times

giqantumhub 1 xraybmd/nc-bmd-cpu **子** Open Client Welcome back, xraybmd File Kernel Tabs Settings Help **Fdit** Run **☑** Launcher \times M hip_bmd.ipynb \mathcal{C} 日 + ※ □ □ ■ C >> Markdown v Python 3 \bigcirc α Filter files by name \blacksquare / code / Create the inference engine Name **Last Modified** untracked an hour ago $[1]$: import os hip_bmd.ipynb an hour ago import sys os.chdir('/mnt/labbook/input/bmd-model') spine_bmd.ipynb an hour ago # sys.path.append('/mnt/Labbook/input/bmd-model') import logging import logging handlers import json Step 10: Now you can run the notebooks for hip and import glob import shutil import flask spine BMD estimation in the client. We provide 100 import yaml import copy test images for hip and spine. You can change the from PIL import Image, ImageDraw, ImageFont import numpy as np from easydict import EasyDict as edict variable "idx" to select the image for testing. After from racad.utils.dcm2png import dicom2img_itk from racad.utils import status_code running, a visualization of the landmark detection, from dag. Scripts. infer landmark import InferenceEngine as Da quality assessment and estimated BMD will be display from racad.tester.infer_hip_bmd import InferenceEngine as Bm from racad.scripts.hip_bmd.prepare_roi_data import extract_h in the notebook. from trauma. PXR. inference. engine import PxrInferenceEngine, # Model directories hip_bmd_dir = './hip_bmd/bmd_estimator' Please note that the code runs on Gigantum clients pelvis_lmk_dir = './hip_bmd/pelvis_detector' pxr_fx_dir = './hip_bmd/fx_detector' hip_lmk_dir = './hip_bmd/hip_detector' without GPU support. So the speed is slow. It runs pxr_implant_dir = './hip_bmd/implant_detector' # Initialize pelvis fracture detector much faster with GPU. with $open(pxr_fx_dir + "/run.yaml")$ as f: config = edict(yaml.load(f, Loader=yaml.FullLoader)) $pelvix_fx_{\text{engine}} = PyrInference Engine (config)$ # Initialize pelvis implant detector with open(pxr_implant_dir + '/run.yaml') as f : config = edict(yaml.load(f, Loader=yaml.FullLoader)) nelvis imnlant engine - DyrImnlantInferenceFngine(config)

☆ ■ ◆

 \Box