Project Codebook

Variable Name	Label/Description	Variable Codes	Variable Format	Value Indicating Missing Data
StudyInstanceUiD	Study ID for each patient scan (ID corresponding to folders containing multiple images)	1.2.826.0.1.3680043.[int_no_of_images]	Float	None
patient_overall	Whether the patient has a fracture	0 = No Fracture Overall 1 = Has Fracture	Binary: numeric	None
C1	Whether C1 (first vertebrae) has a fracture	0 = no fracture 1 = has fracture	Binary: numeric	None
C2	Whether C2 has a fracture	0 = no fracture 1 = has fracture	Binary: numeric	None
C3	Whether C3 has a fracture	0 = no fracture 1 = has fracture	Binary: numeric	None
C4	Whether C4 has a fracture	0 = no fracture 1 = has fracture	Binary: numeric	None
C5	Whether C5 has a fracture	0 = no fracture 1 = has fracture	Binary: numeric	None
C6	Whether C6 has a fracture	0 = no fracture 1 = has fracture	Binary: numeric	None
C7	Whether C7 (last vertebrae) has a fracture	0 = no fracture 1 = has fracture	Binary: numeric	None

Notes:

1. We are working with image data in multiple folders, the variable names do not correspond to each individual image. Instead, they correspond to all the images in a given folder. For example if C1 = 1 and C2 = 0 for folder 1, it means that all the images in folder 1 have a fracture on C1 and no fracture on C2.

2. There are 2019 image folders and about 713k image files from the training dataset.
3. There is not a single folder that is empty and so we don’t have missing data.