



## Dataset Documentation

MURA v1.1 — Musculoskeletal Radiographs · v1.0

■ <https://www.kaggle.com/datasets/cjinny/mura-v11>

### 1. Dataset Overview

Name	MURA v1.1 — MUsculoskeletal RAdiographs
Source	Stanford ML Group — Kaggle
URL	<a href="https://www.kaggle.com/datasets/cjinny/mura-v11">https://www.kaggle.com/datasets/cjinny/mura-v11</a>
Task	Multi-Class Classification — Body Part + Abnormality Detection
Domain	Medical X-ray Imaging — Musculoskeletal Radiology
Format	PNG images — RGB channels — 320x320 px (after resize)
Split	Train / Validation (15% stratified from train) / Test (valid folder)

### 2. Dataset Statistics

**~36,808**  
Training Images  
After 85/15 split

**~6,496**  
Validation Images  
15% stratified

**~3,197**  
Test Images  
MURA valid folder

**14**  
Classes  
7 regions × 2

### 3. Body Regions & Class Distribution

Body Region

Test Samples

Test Accuracy

Notes

ELBOW	465	83.4%	Best performing region
WRIST	659	83.0%	Largest test set
HUMERUS	288	78.5%	Smallest test set
HAND	460	76.7%	High Normal/Abnormal confusion
SHOULDER	563	76.0%	Complex anatomy
FINGER	461	75.5%	Small bone detail
FOREARM	301	70.4%	Lowest accuracy

## 4. Class Label Mapping — 14 Classes

Label	Class Name	Label	Class Name
0	ELBOW_Normal	1	ELBOW_Abnormal
2	FINGER_Normal	3	FINGER_Abnormal
4	FOREARM_Normal	5	FOREARM_Abnormal
6	HAND_Normal	7	HAND_Abnormal
8	HUMERUS_Normal	9	HUMERUS_Abnormal
10	SHOULDER_Normal	11	SHOULDER_Abnormal
12	WRIST_Normal	13	WRIST_Abnormal

## 5. Dataset Challenges

<b>Class Imbalance</b>	Unequal sample counts across body regions — addressed with balanced class weights.
<b>High Inter-Class Similarity</b>	Normal vs Abnormal distinction can be subtle — addressed with Focal Loss.
<b>Low Contrast Images</b>	X-rays have inherently low contrast — addressed with contrast augmentation.

**Multiple Anatomical Regions**

Single model must classify 7 different body parts simultaneously.

**Pathology Subtlety**

Hairline fractures and minor abnormalities may appear nearly identical to normal.

## 6. Citation

Rajpurkar, P., Irvin, J., Ball, R. L., et al. (2018). MURA: Large Dataset for Abnormality Detection in Musculoskeletal Radiographs. *arXiv preprint arXiv:1712.06957*. | Stanford ML Group.

■■ *Dataset used for research and educational purposes only — not clinical use.*