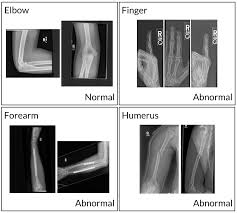
**Uma Maheshwari T M**

Email : umamohantm@gmail.com

Mobile : +91 9952556206

**Abnormality Detection in Bone X-Ray Radiographs**

**September, 2020**



**Overview**

Bio-medical image analysis is an interdisciplinary field which includes: biology, physics, medicine and engineering. It deals with application of image processing techniques to biological or medical problems. Medical images to be analyzed contain a lot of information regarding the anatomical structure under investigation to reveal valid diagnosis and thereby helping doctors to choose adequate therapy. Doctors usually analyze these medical images manually through visual interpretation.

**Goals**

Diseases and injuries to the bone are the major contributing factors in causing abnormalities in bones. Whenever there is an injury to the bone, the physician asks you to do an X-Ray, thus when such hundreds of patients visit hospitals everyday there are massive numbers of X-rays done on a regular basis. To be specific with the stats, musculoskeletal conditions affect more than 1.7 billion people worldwide, and are the most common cause of severe, long-term pain and disability, with 30 million emergency department visits annually and increasing. So, in order to reduce the error rate of the Radiologist and to do the analysis much faster, an AI solution should suffice the purpose.

**Data Set**

Dataset used in this project is called MURA. MURA is one of the largest public radiographic image datasets. MURA is a dataset of musculoskeletal radiographs consisting of 14,863 studies from 12,173 patients, with a total of 40,561 multi-view radiographic images. Each belongs to one of seven standard upper extremity radiographic study types: elbow, finger, forearm, hand, humerus, shoulder, and wrist. You can download the dataset from the official contest website here. <https://stanfordmlgroup.github.io/competitions/mura/>.

This dataset contains four CSV files, two files for training the model and other two for validation. This CSV files contains the path of images stored in the directory and the labels.

**Data Wrangling:**

Data wrangling is the second step in Data Science methods which makes the data more suitable for further analysis. It includes four steps

* Data Collection.
* Data Organization.
* Data Definition.
* Data Cleaning.

**1. Data Collection:** Musculoskeletal conditions affect more than 1.7 billion people worldwide, and are the most common cause of severe, long-term pain and disability; with 30 million emergency department visits annually and increasing. MURA is a dataset of musculoskeletal radiographs consisting of 14,863 studies from 12,173 patients, with a total of 40,561 multi-view radiographic images. Each belongs to one of seven standard upper extremity radiographic study types: elbow, finger, forearm, hand, humerus, shoulder, and wrist. Each study was manually labeled as normal or abnormal by board-certified radiologists from the Stanford Hospital at the time of clinical radiographic interpretation in the diagnostic radiology environment between 2001 and 2012. Stanford ML group made this dataset available to all and hosted a Bone X-Ray Deep Learning competition.

**2. Data Organization:** This step of Data Science involves creating sub folders to ensure the project is in well organized manner. Here folders data, figures and models are created to hold the outputs of further steps of the project. And this step includes building a well defined project structure in Github.

Github link to this capstone project: <https://github.com/umamohantm/springboard/tree/master/Capstone%202>

**3. Data Definition:** Data Definition includes defining the data such as column names, data types of the column, description of the column, count of unique values or codes and range of unique values or codes including NAN values.

**4. Data Cleaning:** The data originally obtained was in CSV file and directly loaded into the pandas data frame effortlessly. In other words, the dataset we have in our hands is already relatively clean. So no data cleaning is required here.

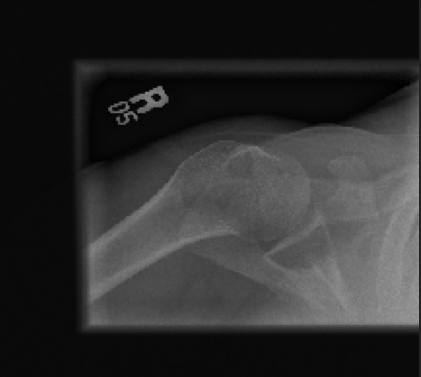
**Exploratory Data Analysis**

**This section involves the visualization part to know the insights from the dataset.**

**Visualizing abnormal bone X-Ray Radiographs from the dataset**

****

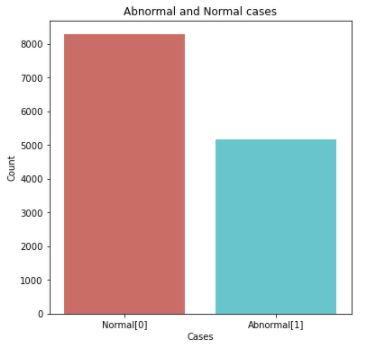
**Visulaising normal bone X-Ray Radiographs from the dataset**

****

**Checking whether the data frame have Imbalanced data**

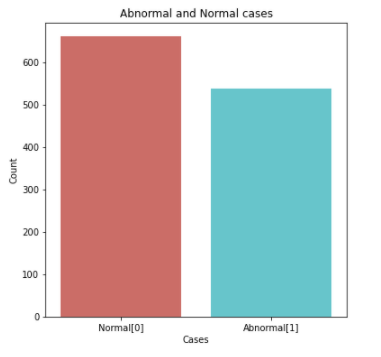
First checking with train\_labelled dataframe which has two column train\_path and train\_labels. This path contains information about individual patient not for individual images.

**Abnormal and normal cases count in train\_labelled DataFrame**



* Almost the train\_labelled data frame has balanced data 8000 in Class Normal [0] and 5000 in class Normal [1].
* train\_labelled DataFrame is not labeled for each image. Actually the images are grouped for one patient in the dataset.
* Since it is not been labeled separately for each image, it is better to consider train\_image as training dataset.

**Abnormal and normal cases count in valid\_labelled DataFrame**

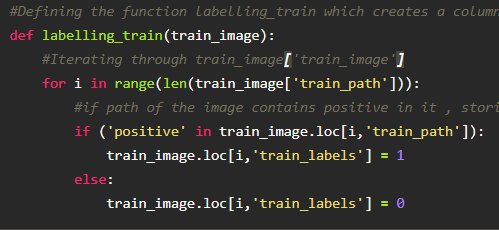
******

* This Dataframe contains almost balanced data 600 in Class Normal [0] and 550 in class Normal [1].
* valid\_labelled Dataframe is not labeled for each image. Actually the images are grouped for one patient in this dataset same as train\_labelled Dataframe. Since this Dataframe doesn’t contain information about images separately we can consider valid\_image Dataframe for validating the model.

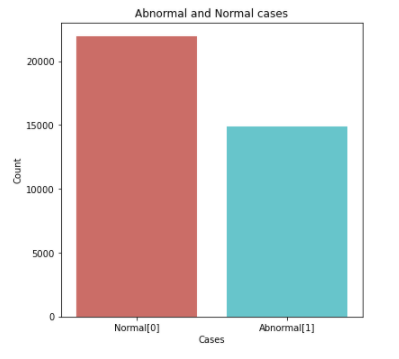
**Abnormal and Normal cases count in train\_image Dataframe**

This training Dataframe doesn’t contain label information column separately, so need to extract that from the path available and create a column train\_labels.

**Code:**

**

**Bar Plot**

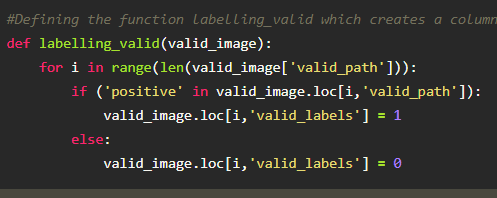
******

The training Dataframe contains balanced sets of abnormal and normal X-Ray images. It is not an imbalanced dataset.

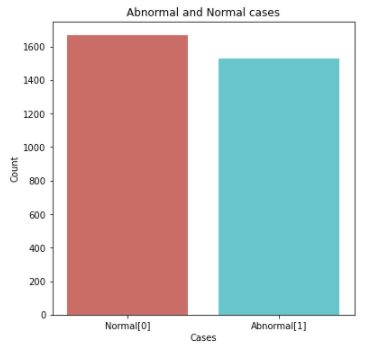
**Abnormal and normal cases count in valid\_image Dataframe**

This validating Dataframe doesn’t contain label information column separately, so need to extract that from the path available and create a column valid\_labels.

**Code:**

******

**Bar Plot**

******

The validating Dataframe contains balanced sets of abnormal and normal X-Ray images. It is not an imbalanced dataset

**Generating body parts and study type columns for train\_image dataframe**

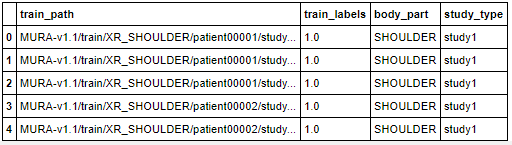
The training dataset contains only the path and labels columns. From the path column we generate body parts and study type columns

**Code:**

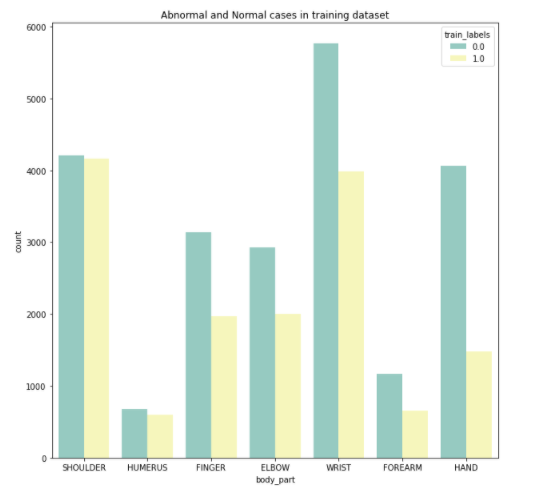
Below is the function used to generate body\_part and study\_type column for training dataset



After generating the body\_part and study\_type columns the training dataframe look like this.



**Abnormal and Normal Cases in training dataset**

******

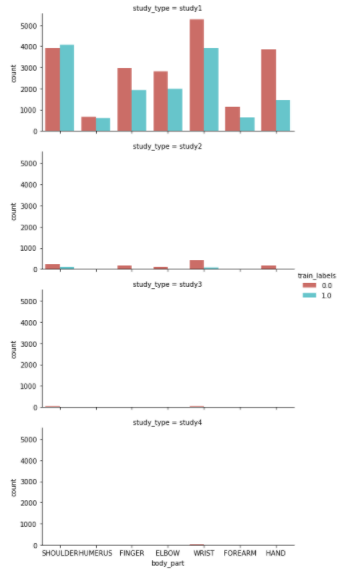
This training dataset contains X-rays of following body-parts

* Shoulder
* Humerus
* Finger
* Elbow
* Wrist
* Forearm
* Hand

Even if we consider the body-parts separately, we could observe that the data is almost balanced.

**Abnormal and normal cases in training dataset hued by studytype column**

**Plotting against body\_part and count of X-rays (training dataset)**

****

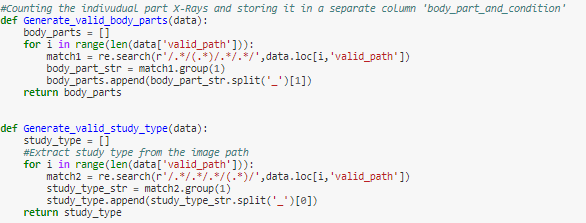
In training dataset if we plot against body\_part and the count of x-rays hued by study\_table, we could see that more data is under study\_table 1 and study\_table 2 stand next to it.

**Generating body parts and study type columns for valid\_image dataframe**

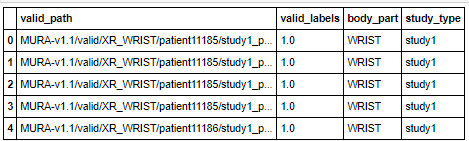
The validating dataset contains only the path and labels columns. From the path column we generate body parts and study type columns

**Code:**

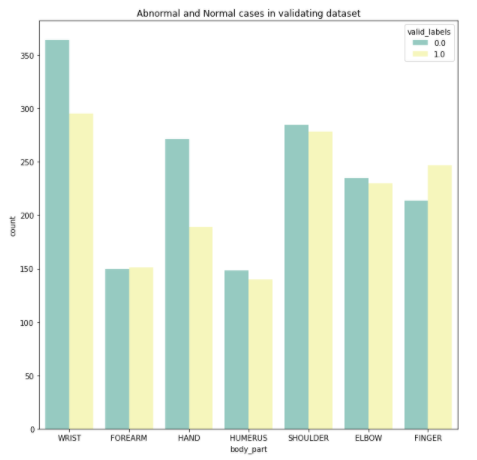
Below is the function used to generate body\_part and study\_type column for validating dataset



After generating the body\_part and study\_type columns the validating dataframe look like this.



**Abnormal and Normal Cases in training dataset**

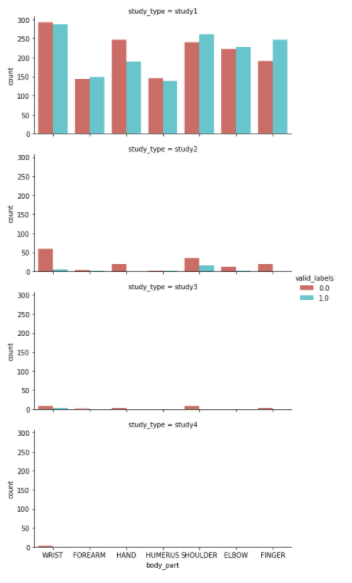
****

This validating dataset contains X-rays of following body-parts

* Wrist
* Forearm
* Hand
* Humerus
* Shoulder
* Elbow
* finger

Even if we consider the body-parts separately, we could observe that the data is almost balanced.

**Plotting against body\_part and count of X-rays (validating dataset)**

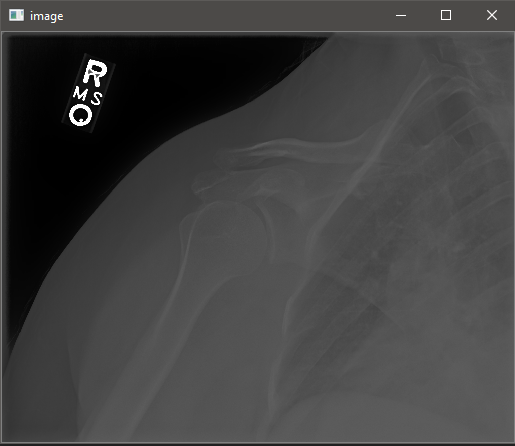
**

In validating dataset if we plot against body\_part and the count of x-rays hued by study\_table, we could see that more data is under study\_table 1 and study\_table 2 stand next to it. The same is observed in training dataframe also.

**Preprocessing**

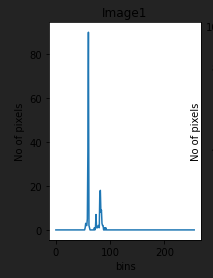
In Preprocessing the image enhancements are done using openCV, Keras, scikit libraries

**Original Image**



The enhancement methodoligies used are

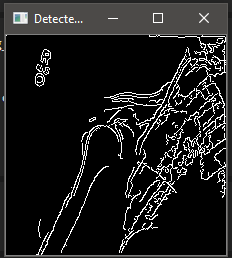
* CLAHE (Contrast Limited Adaptive Histogram Equliser)
* cv2.canny()
* cv2.GaussianBlur
* unsharp\_mask() from scikit library.
* Considering an X-Ray image from the training dataset the image enhancement used in this project is explained below,
* The image contrast is increased using Contrast Limited Adaptive Histogram Equliser (CLAHE). The histogram of original image is below,



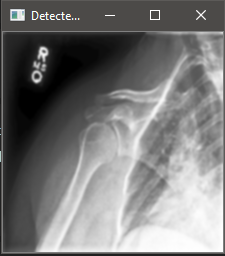
* After applying CLAHE



* For detecting the edges cv2.canny() is used. After detecting the edges we can crop the original image based on the dimensions obtained by detecting the edges of the original image.



* For blurring and smoothing the image cv2.GaussianBlur() is used.



**Before and After image enhancements**

