Identifying Bone X-rays
Adeyemi Adejuwon

Capstone 2

Springboard Data Science

September 2020

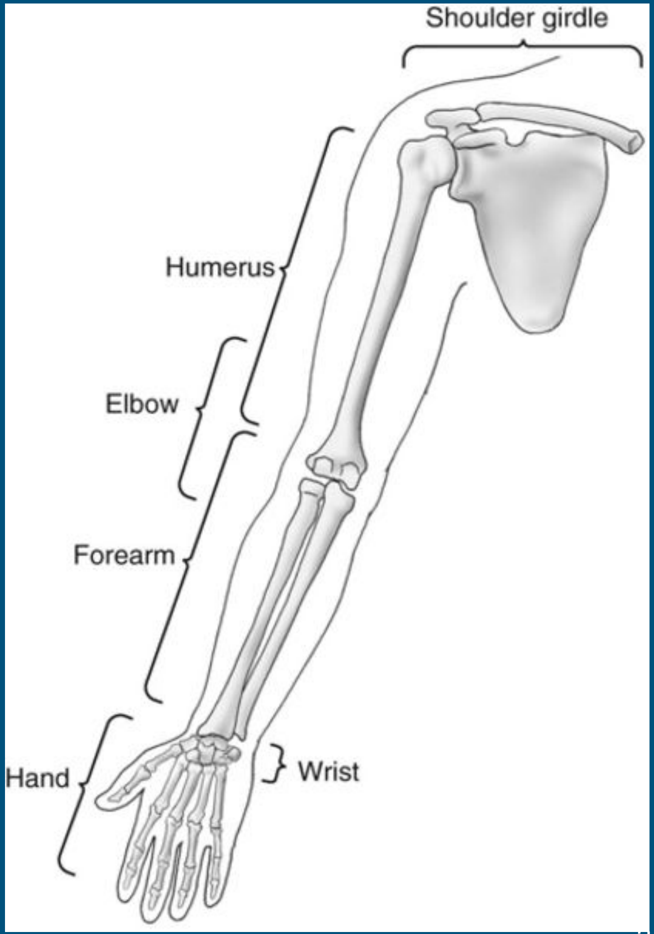
Description of Dataset

●MURA (**MU**sculoskeletal **ra**diographs) is a large dataset of bone X-rays supplied by Stanford.

●Each study contains one or more images and is manually labelled by radiologists as ‘Normal’ or ‘Abnormal’.

●The images are X-rays of different body parts-wrist, shoulder, elbow, hand, finger, forearm and humerus.

●MURA dataset comes with train and valid folders containing corresponding datasets



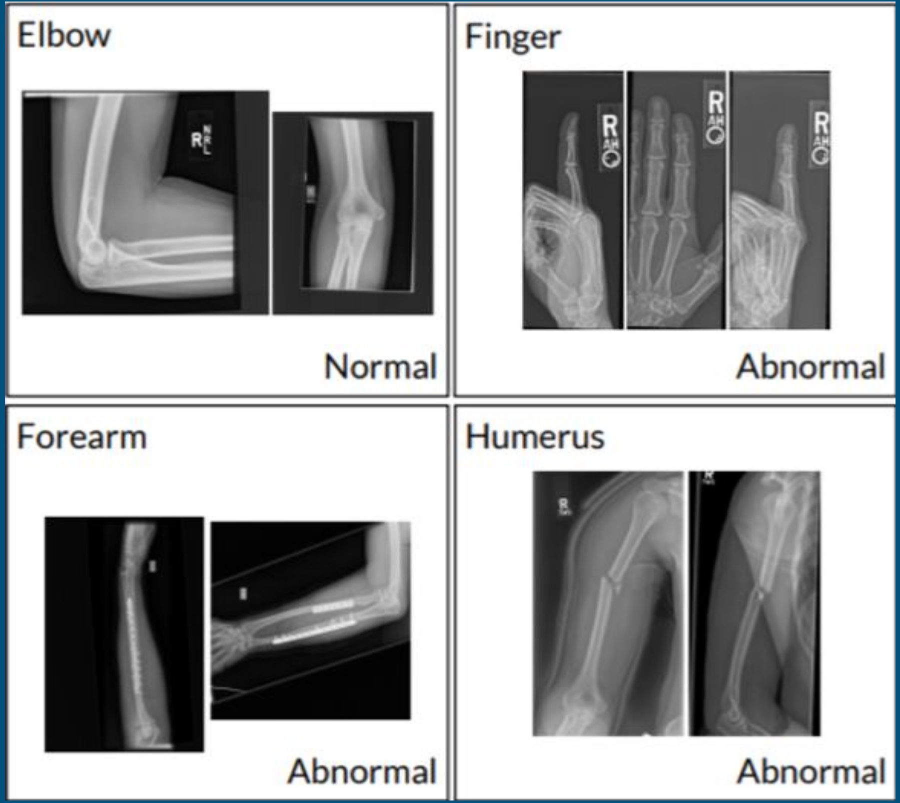
Objective

Classify radiographic images of upper extremity body parts as Normal or Abnormal using machine learning models

The machine learning models used are:

1. **Logistic Regression**
2. **K Nearest Neighbor**
3. **SVM**
4. **Random Forest**

Summary of Dataset



- ~41,000 images from ~15,000 studies (patients) including train and validation sets

- 9,000 studies of normal (negative condition) and 6,000 of abnormal studies (positive condition)

- Typical image res: 500 x 500 pixels

Distribution of the normal (0 class) vs abnormal (1 class) X-rays in our dataset

Fig 7: Number of patients versus body parts for both Normal and Abnormal Data for Training data

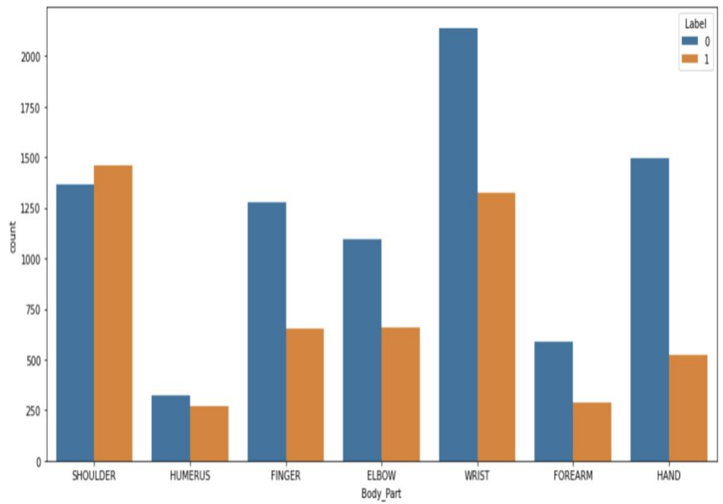
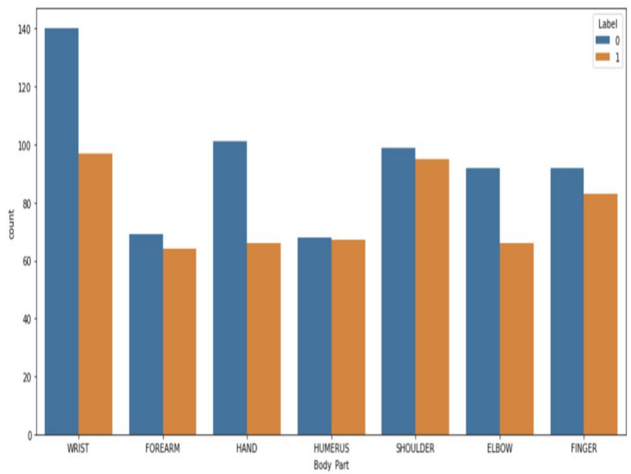
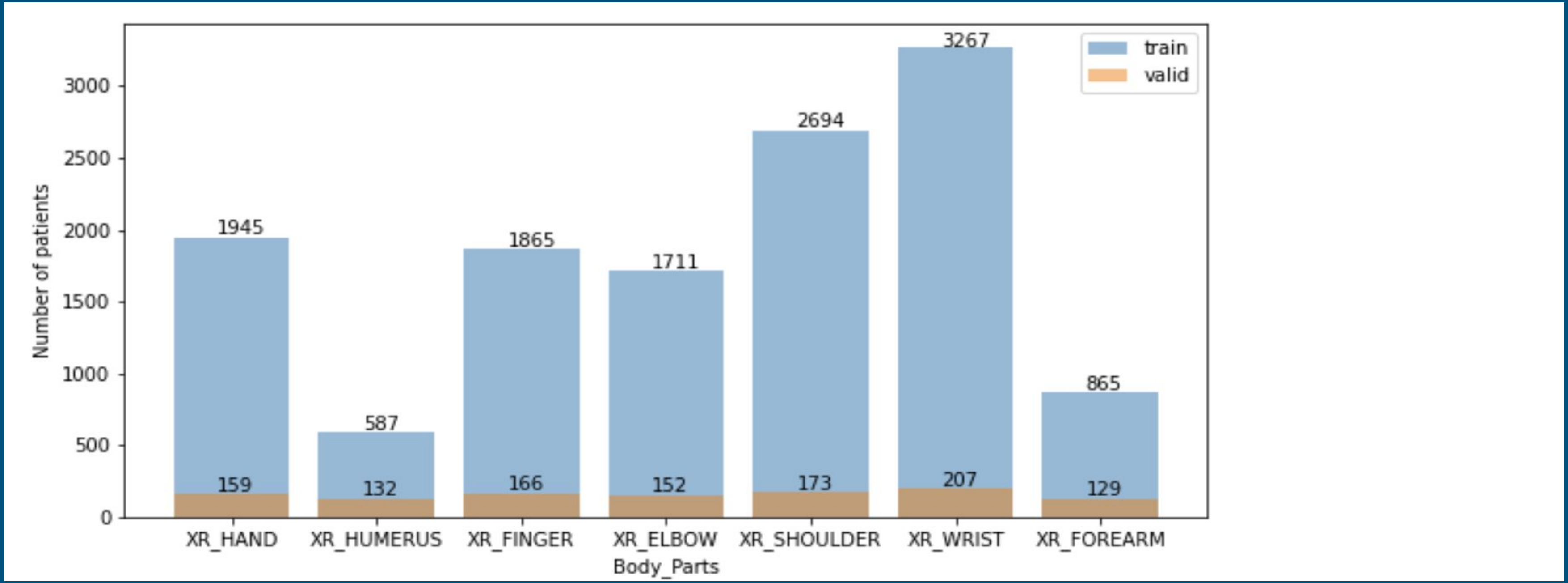


Fig 8: Number of patients versus body parts for both Normal and Abnormal Data for Test data



Distribution of X-rays versus the number of patients in the dataset



Data Wrangling

Preprocessing of Images

●Resizing the images to 224 x 224 to enable the use in the machine learning model

● Normalization of images (scale:0-1) to ensure each input parameter has a similar data distribution

Feature Extraction

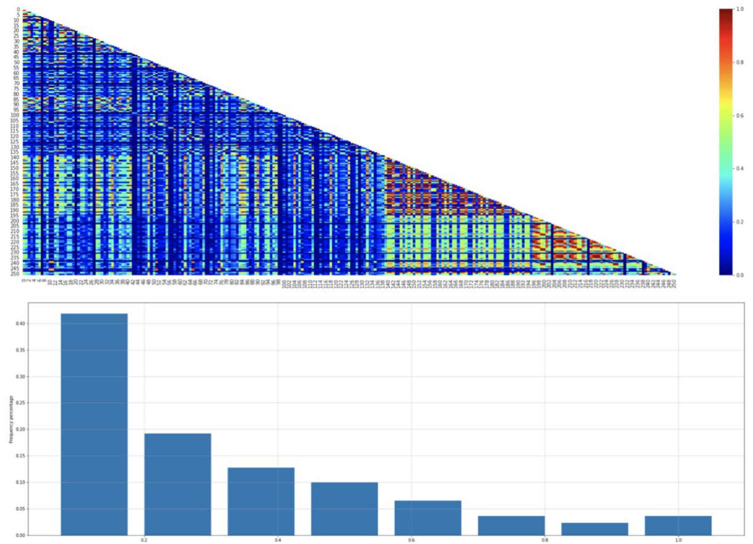
Feature extraction was done using the Gray Level Co-occurrence Matrix (GLCM) method

Results: 252 features for each X-ray image

- 14 features from Texture,
- 14 features from FFT,
- 56 features from GLCM,
- 56 features from GLDM, and
- 112 features from Wavelet

Inferential Statistics

Fig. 11. Heat map and Histogram representation of Pearson correlation Coefficients for the Humerus



- Feature Extraction done with GLCM
- 85% of the features have correlation coefficients of less than 0.5

Machine Learning Models for Abnormal Class

Model	Logistic Regression			Random Forest		
Body Type	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Forearm	0.67	0.12	0.21	0.74	0.22	0.34
Wrist	0.43	0.19	0.26	0.7	0.4	0.51
Finger	0.76	0.46	0.57	0.78	0.59	0.67*
Hand	0.6	0.05	0.08	1.00	0.09	0.17
Humerus	0.56	0.6	0.58	0.69	0.54	0.61
Shoulder	0.5	0.58	0.54	0.62	0.69	0.65
Elbow	0.62	0.12	0.2	0.87	0.2	0.32

*Best F1-Score is for Finger and Random Forest

Machine Learning Models for Abnormal Class

Model	SVM			KNN		
Body Type	Precision	Recall	F1-Score	Precision	Recall	F1-Score
Forearm	0.73	0.12	0.21	0.59	0.36	0.45
Wrist	0.57	0.32	0.41	0.59	0.48	0.53
Finger	0.81	0.46	0.58	0.68	0.39	0.49
Hand	0.5	0.02	0.03	0.62	0.15	0.24
Humerus	0.65	0.64	0.65	0.63	0.69	0.66
Shoulder	0.54	0.62	0.58	0.58	0.69	0.61
Elbow	0.62	0.12	0.2	0.69	0.27	0.39

Discussion of Results

- The best F1 score is generated by the Random Forest model for finger x-rays
- Classification report for the abnormal class shows that :
  a. All models offer poor F1-scores for the elbow, forearm, hand and wrist
  b. All models offer positive F1-scores for the finger, humerus and shoulder
-

Conclusion

- The performance evaluation of the abnormality detection in the MURA dataset was performed using three statistical parameters such as recall, precision and F1 score.

- Random Forest for the X-ray images of **fingers** provides the best performance metrics when predicting the abnormal class for the upper extremity body part.

Recommendations for the Client

- Random Forest is the model of choice when trying to predict X-ray images for shoulder, humerus and fingers given that F1-scores for these body parts are greater than 0.6.

- Radiologist classify images when having to interpret the wrist, hand, forearm and elbow, rather than using the machine learning models to assist his predictions. This is because F1-scores are less than 0.5 for these body parts.

Future & Extension of Project

1. Use deep learning algorithms such as CNN to see if accuracy of model classification can be improved.

2. Include more pre-processing steps with the images. These pre-processing steps could include adding masks to the images, and applying different transforms, to make the currently extremely varied data more uniform in contrast, orientation, and scale.

3. Develop a digital data repository for radiologists to upload data securely. Results from these X-rays are then interpreted with a Random Forest model to see if X-rays are normal or abnormal.

Works Cited

1. "What Is MURA?" *MURA Dataset: Towards Radiologist-Level Abnormality Detection in Musculoskeletal Radiographs*, stanfordmlgroup.github.io/competitions/mura/.

2. Khuzani, Abolfazl Zargari, et al. "COVID-Classifier: An Automated Machine Learning Model to Assist in the Diagnosis of COVID-19 Infection in Chest x-Ray Images." *MedRxiv : the Preprint Server for Health Sciences*, Cold Spring Harbor Laboratory, 18 May 2020, www.ncbi.nlm.nih.gov/pmc/articles/PMC7273278/#S6.