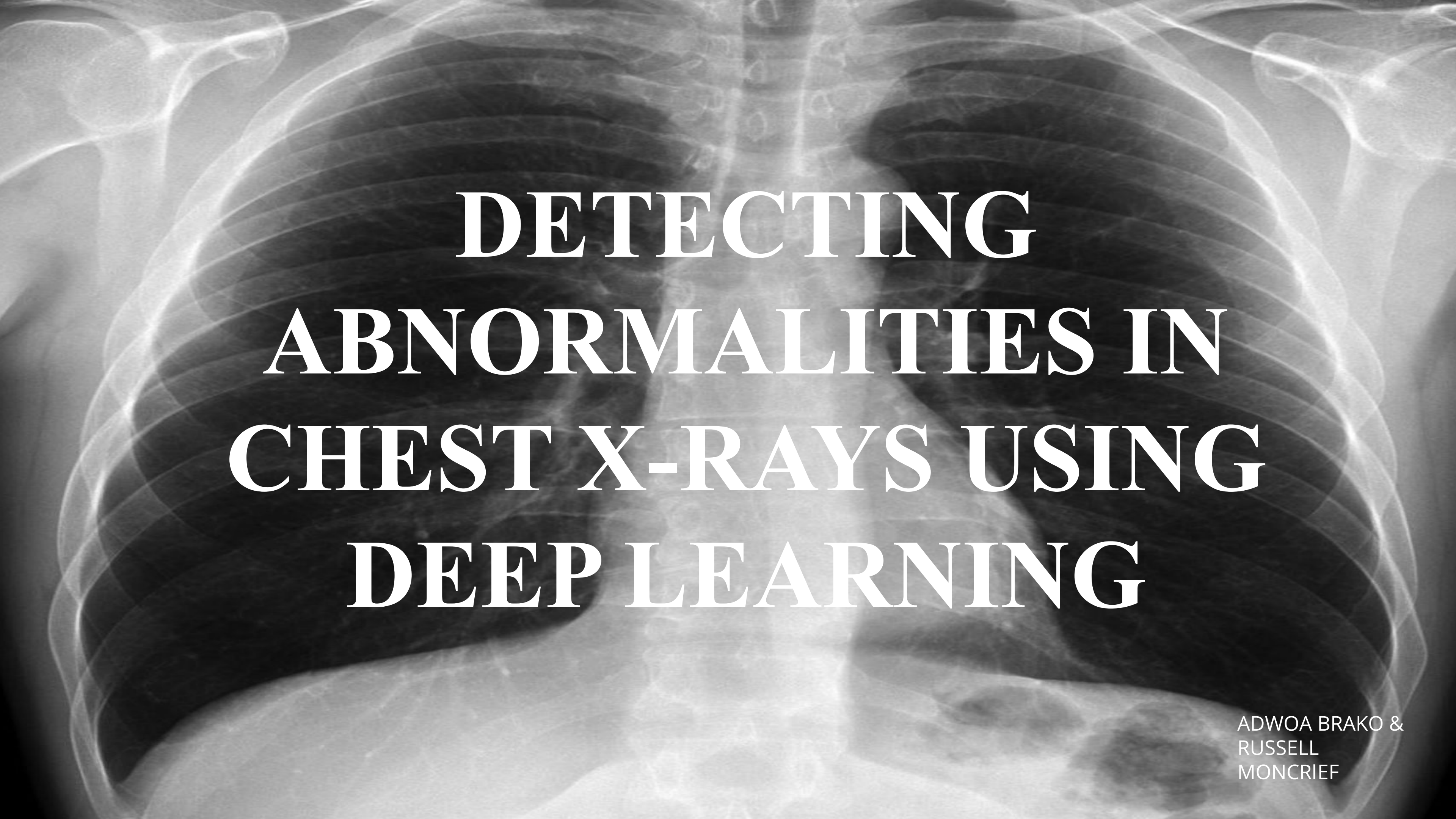


DETECTING ABNORMALITIES IN CHEST X-RAYS USING DEEP LEARNING

ADWOA BRAKO & RUSSELL MONCRIEF

OUTLINE

- Introduction
- Problem Statement & Objective
- Data
- Preliminary Analysis
- Preprocessing
- Model
- Discussion & Limitations
- Conclusion

Introduction

- The number of X-Ray scans conducted in the US continues to steadily increase. Of these scans, a large portion is of chest X-Rays (approximately 45%)

- Functionality of chest X-Rays can be limited by challenges in interpretation, which may lead to delayed and/or wrong diagnosis of diseases

Problem Statement

- Given the high rate of chest X-Ray scans performed, how can radiologists make objective diagnosis, with less errors, within a timely manner?

Objectives

- Develop an algorithm to aid in classifying Chest X-Ray images based on the type of disease

- Accurate classification of these images will help with the early detection of diseases and adequate timely diagnosis

Data

This project is based on the National Institutes of Health Chest-X-ray dataset.

- The dataset is composed of over 100,000 grayscale chest x-ray images (with sizes of 1024 x 1024 pixels) from over 30,000 unique patients

- Labels with 15 different disease classes, such as **Infiltration**, **Fibrosis, Cardiomegaly, Pneumonia, Atelectasis, Nodule,** among others

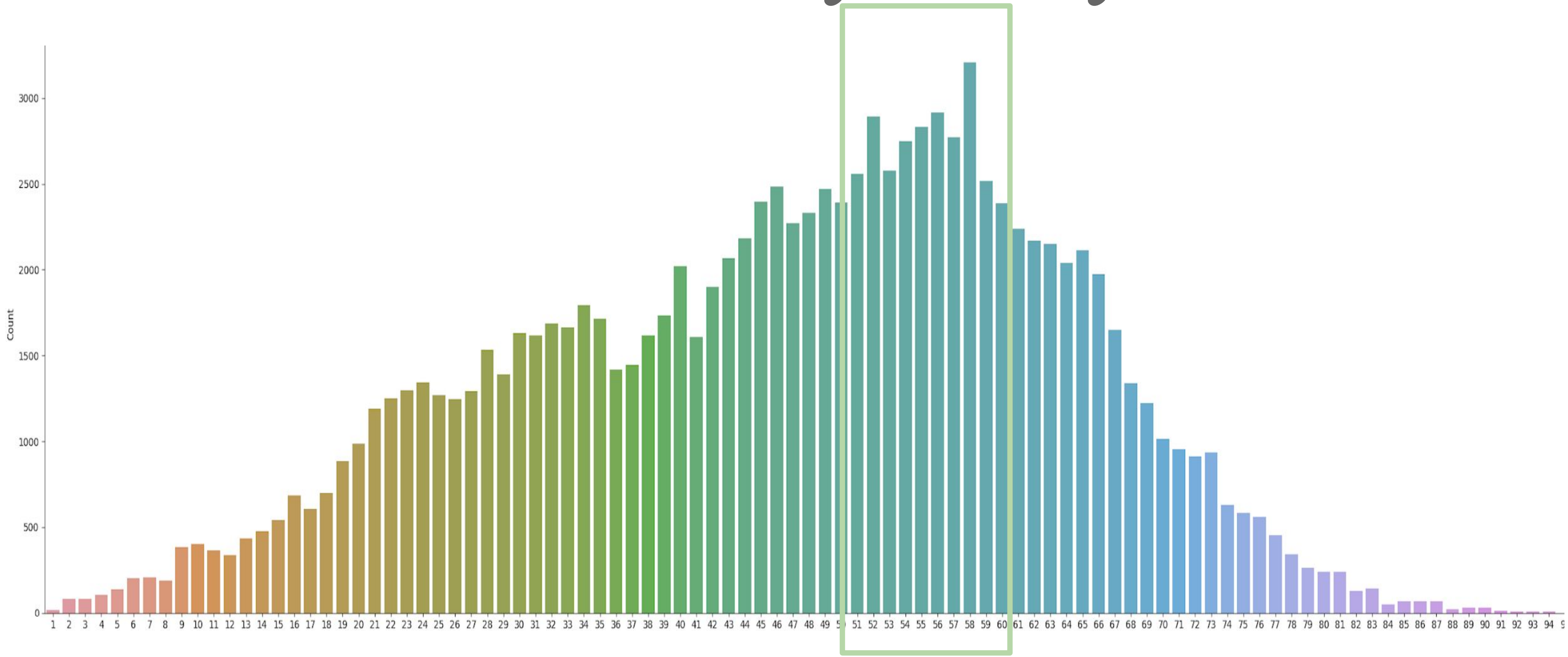
The image labels are produced with Natural Language Processing and associated radiological reports. The labels are thought to be greater than 90% accurate.

DATA CLEANING & PRELIMINARY ANALYSIS

Data Cleaning

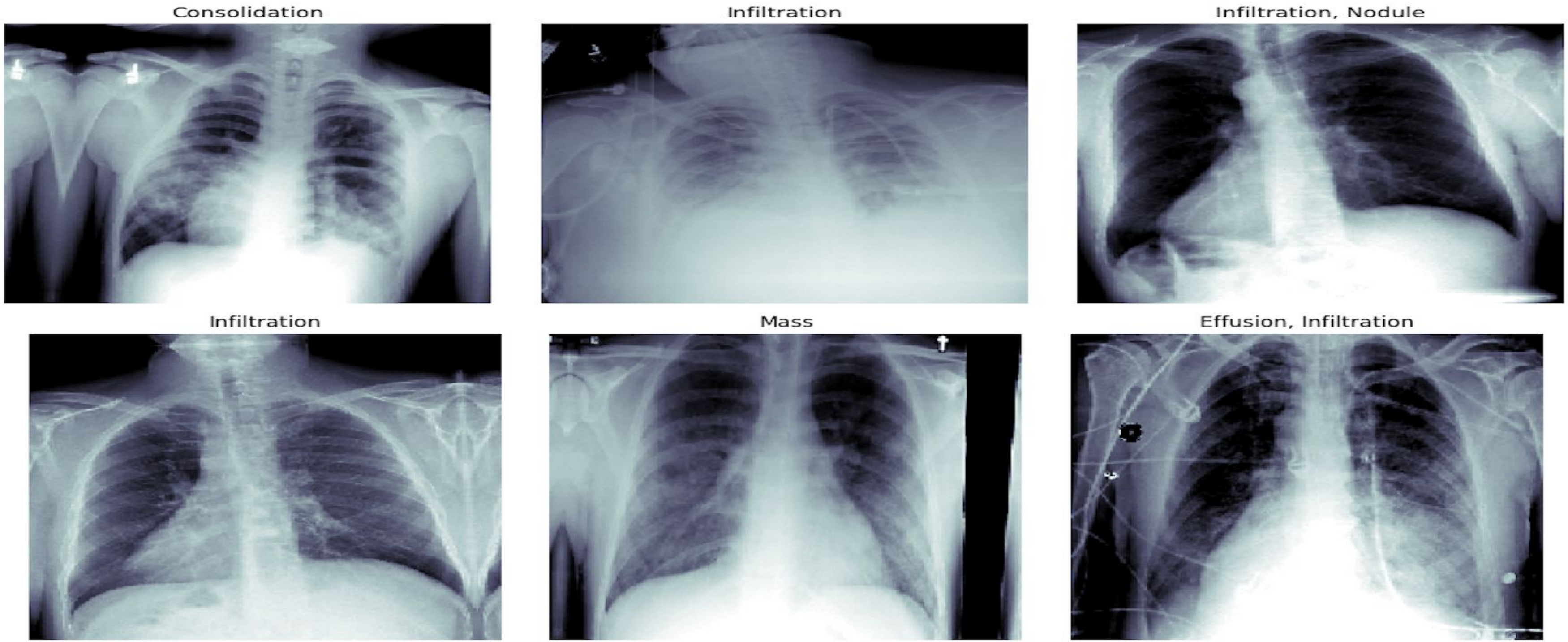
- Drop unnecessary columns in dataframe
- Create new column with image paths;
    - to map images to respective labels
- One-Hot encode target labels

Preliminary Analysis



Age distribution of patients

Preliminary Analysis



PREPROCESSING

Preprocessing

Split Data

Split data into **Train**, **Validation** and **Test** in the ratio (70, 10, 20)% respectively

Resize Images

Reduce image sizes from (1024 x 1024) pixels to (224 x 224) pixels

Implement Keras Framework

Used Keras to transform images and initiate model for image classification

Data Augmentation

Used Keras ImageDataGenerator to expand size of training data, by creating modified versions of input images

MODELS

Model

CNN Model Architecture

- Two Convolution Layers
  - Kernel size of (6 x6) in first layer and (2 x 2) in second Conv2D layer
  - MaxPooling2D of size (2 x 2) in both layers
  - Dropout
  - Dense Layer

- Activations used in layers: Relu, Sigmoid.

Model - Transfer Learning

MobileNet Pre-trained Model

- This was used because of its lightweight architecture and use of depth-wise separable convolutions.
    - The architecture consists of GlobalAveragePooling2D
    - Dropout of (0.5) was added, as well as a Dense layer
    - Model CheckPoint, Early Stopping and ReduceLROnPlateau callbacks were used during training.

Model Performance On Test Set

55% Accuracy

CNN Model

The model was trained with 2 epochs, with 1227 steps per epoch

- Test Loss : 0.68

- Dropout - 0.2
- Learning Rate - 0.02

90% Accuracy

CNN Model

The model was trained with 3 epochs, with 1227 steps per epoch

- Test Loss : 0.20

- Dropout - 0.5
- Learning Rate - 0.001

75% Accuracy

MobileNet

The model was trained with 1 epoch, with 1227 steps per epoch

- Test Loss : 0.27

- Dropout - 0.5
- Learning Rate - 0.02

92% Accuracy

MobileNet

The model was trained with 3 epochs, with 1227 steps per epoch

- Test Loss : 0.19

- Dropout - 0.5
- Learning Rate - 0.02

Discussion & Limitations

Discussion

- Preprocessing can be enhanced
  - Address over representation of class label “No Finding”
  - Adjust image sizes and image data generation

- More training needs to be done, to improve performance of models
  - Investigate different architectures
  - Improve hyper parameter tuning

Limitations

- Class labels generated through NLP. Not definitive labels.
- Long training time of models

Conclusion

- Preprocessing and Data Augmentation played an integral role in influencing performance of models.
- Models were able to classify Chest X-Ray scans based on labels
- Accuracy of classification can be improved

Contributions