SIIM-FISABIO-RSNA COVID-19 Detection

Using chest radiographs

—Data Science Capstone Project Presentation

Shixin Li
07/03/2021

Contents:

1. Problem Statement
2. Data Wrangling
3. Exploratory Data Analysis
4. Data Pre-processing
5. Modeling
6. Conclusion

Problem Statement

About this Project

1. **What?**

   Covid-19 chest radiographs identification using AI

2. **Why?**

   Help radiologists to diagnose millions of COVID-19 patients more confidently and quickly

3. **How?**

   To develop the supervised deep learning model to classify the image data

About this project

4. Source of the Data

The Foundation for the Promotion of Health and Biomedical Research of Valencia Region(FISABIO), whose primary purpose is to encourage, to promote and to develop scientific and technical health and biomedical research in Valencia Region

5. Criteria of Project Success

Accuracy >= 90%

F1-Score >= 0.80

Data Wrangling

Data Information

Three Data Sets:

1. **Image Data Set: including 6334 one channel chest radiographs in DCM format**
2. **train_image_level.csv**
3. **train_study_level.csv**

```
RangeIndex: 6334 entries, 0 to 6333
Data columns (total 4 columns):
 #   Column            Non-Null Count  Dtype
---  ------            --------------  -----
 0   id                6334 non-null   object
 1   boxes             4294 non-null   object
 2   label             6334 non-null   object
 3   StudyInstanceUID  6334 non-null   object
dtypes: object(4)
memory usage: 198.1+ KB
```

```
RangeIndex: 6054 entries, 0 to 6053
Data columns (total 5 columns):
 #   Column                     Non-Null Count  Dtype
---  ------                     --------------  -----
 0   id                         6054 non-null   object
 1   Negative for Pneumonia     6054 non-null   int64
 2   Typical Appearance         6054 non-null   int64
 3   Indeterminate Appearance   6054 non-null   int64
 4   Atypical Appearance        6054 non-null   int64
dtypes: int64(4), object(1)
memory usage: 236.6+ KB
```

Relationship between the data sets:

1. StudyInstanceUID are corresponding to id in train_study_level.csv
2. Image names are corresponding to the id in train_image_leve.csv

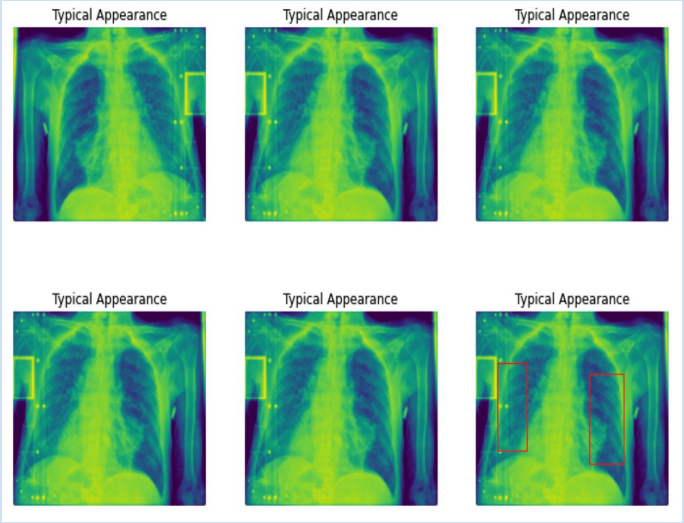
Missing Value: there are 2040 missing values in column ‘boxes’ in train_image_level.csv, but we don’t need to care about it in this project.

Data Cleaning

- **Some study level predictions have multiple image level predictions**

```
5822 of the study level predictions have 1 image level predictions.
207 of the study level predictions have 2 image level predictions.
15 of the study level predictions have 3 image level predictions.
4 of the study level predictions have 4 image level predictions.
3 of the study level predictions have 5 image level predictions.
1 of the study level predictions have 9 image level predictions.
1 of the study level predictions have 6 image level predictions.
1 of the study level predictions have 7 image level predictions.
```

- **In the same study level prediction with more than one image level, the images are the same (see the figure on the right)**
  - remove the duplicated images
  - After removing the images, we have 6054 images left, the same as the number of study level observations.



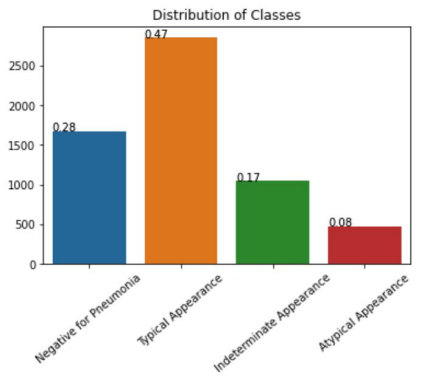
Exploratory Data Analysis

EDA

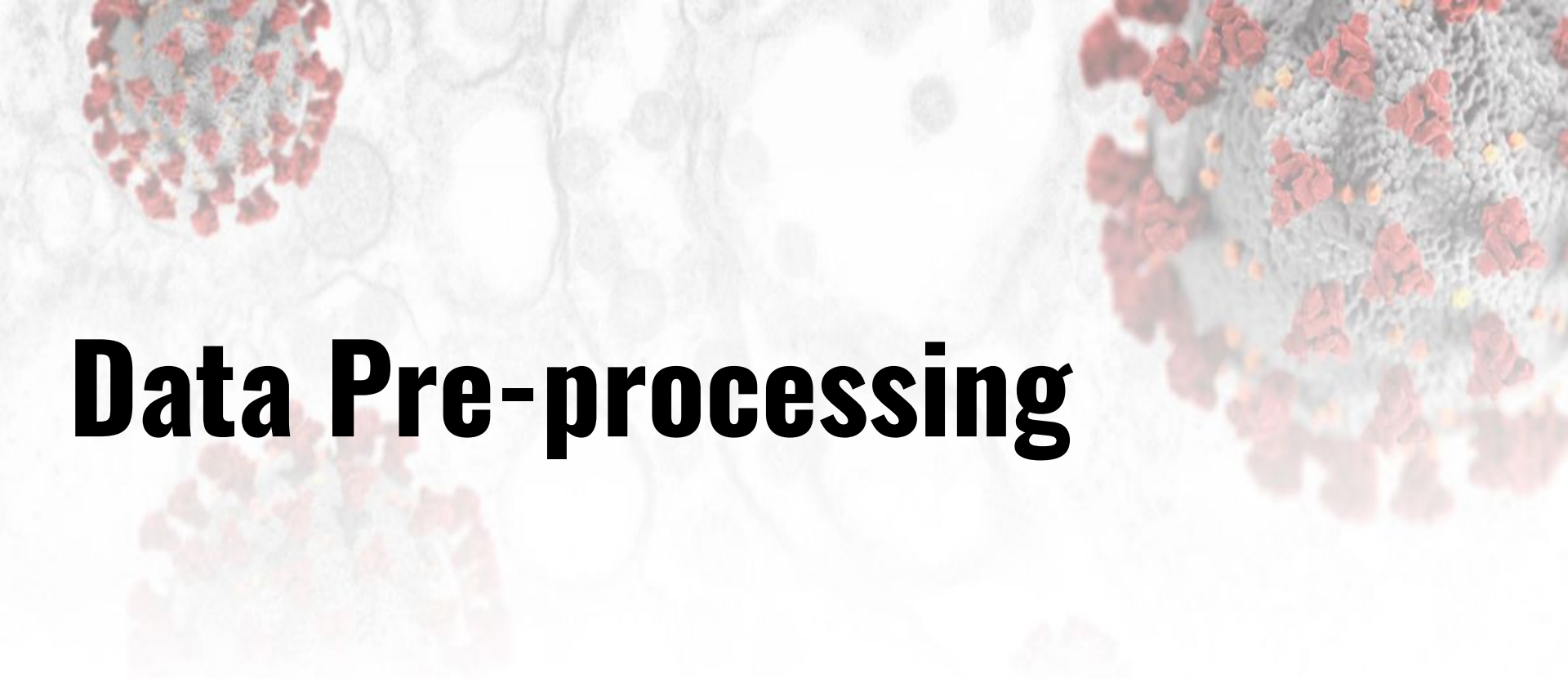
1. There 4 classes images:

- **Typical Appearance:** Multifocal bilateral, peripheral opacities with rounded morphology, lower lung–predominant distribution
- **Indeterminate Appearance:** Absence of typical findings AND unilateral, central or upper lung predominant distribution
- **Atypical Appearance:** Pneumothorax, pleural effusion, pulmonary edema, lobar consolidation, solitary lung nodule or mass, diffuse tiny nodules, cavity
- **Negative for Pneumonia:** No lung opacities

2. Class Distribution



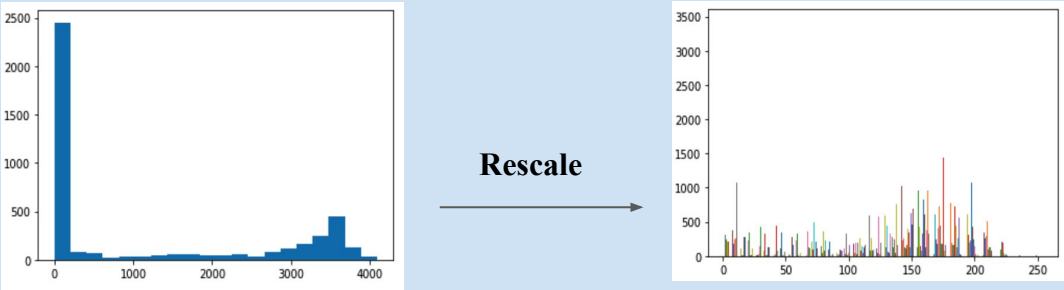
Class Imbalance



Data Pre-processing

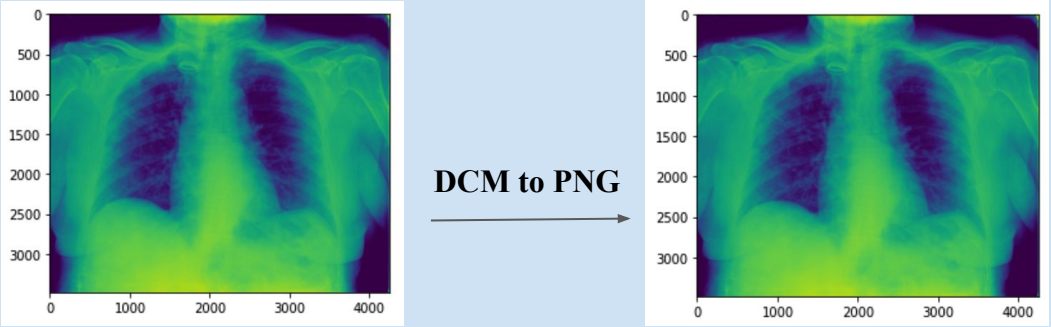
Image Data Pre-processing

- **Rescale the pixel values to range 0-255**



The pixel values of the original images are in a large range, and ranges are different in different images. Rescaling the images can remove some random noise in different images.

- **Transform the image format from DCM to PNG**



Save them as PNG can save a lot of memory and easier for later use. The sample image on the left is rescaled and saved as PNG

Image Data Generating

1. **Image Data Splitting**

   Split the data set into training and validation set in ratio 0.8 and 0.2
2. **Image augmentation**

   Image rotation, horizontal flip, width shift, height shift, shear, zoom in
3. **Image resizing and rescaling**

   Resize the images from different sizes into 224*224

   Rescale the pixel values in range [-1,-1] or [0-1] base on models
4. **Batch size = 64**

Modeling

Models & Metric

Supervised deep learning models:

1. **Vgg16 without pretrained weights;**
2. **Vgg16 with pretrained weights;**
3. **Resnet50 without pretrained weights**

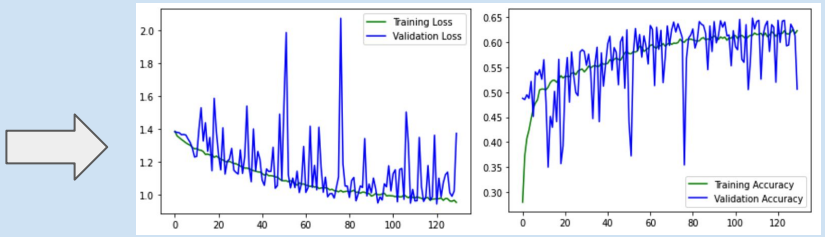
Then, add two fully connected layers at the end of each models with normalization and dropout.

Metric for evaluating the models:

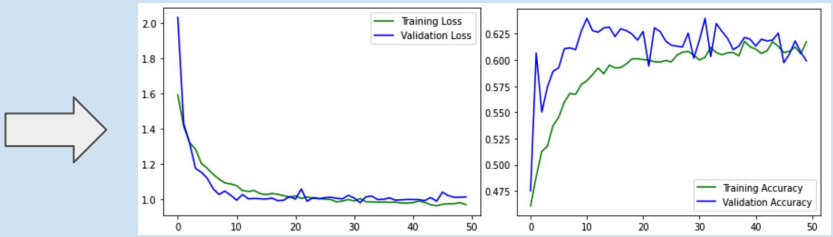
- **Use Area Under the Receiver Operating Characteristics(ROC-ACU) score as the metric to select(or save) the best model during training**
- **Use f1-score as the metric to select the best model from the three different models**

Model Training

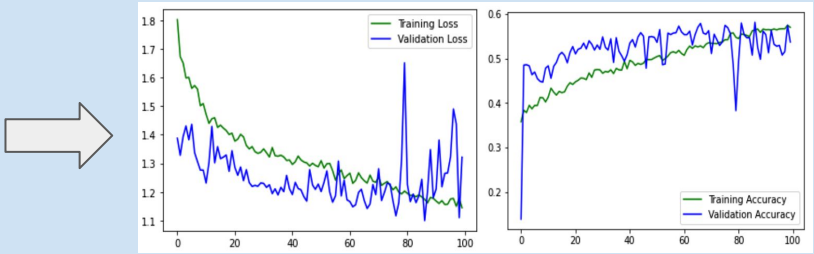
1. **Non pretrained Vgg16, lr = 1e-5**
   - Validation loss and accuracy are very unstable
   - Highest accuracy = 65%
   - Starts overfitting about epoch=90



2. **Pretrained Vgg16, lr = 5e-4**
   - Validation loss and accuracy are more stable
   - Highest accuracy = 64%
   - Validation loss reach the lowest point at epoch=32



3. **Non pretrained Resnet50, lr = 3e-5**
   - Validation loss and accuracy more stable than Non pretrained Vgg16
   - Highest accuracy = 58%
   - Starts overfitting about epoch=75



Model Selection

Testing results from each model with best weights:

Non pretrained Vgg16

	precision	recall	f1-score	support
Atypical Appearance	0.00	0.00	0.00	103
Indeterminate Appearance	0.00	0.00	0.00	168
Negative for Pneumonia	0.62	0.75	0.68	353
Typical Appearance	0.66	0.87	0.75	587
accuracy			0.64	1211
macro avg	0.32	0.41	0.36	1211
weighted avg	0.50	0.64	0.56	1211

Pretrained Vgg16

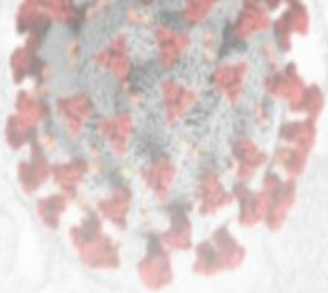
	precision	recall	f1-score	support
Atypical Appearance	0.39	0.11	0.17	102
Indeterminate Appearance	0.19	0.02	0.04	168
Negative for Pneumonia	0.63	0.65	0.64	353
Typical Appearance	0.63	0.86	0.73	587
accuracy			0.62	1210
macro avg	0.46	0.41	0.40	1210
weighted avg	0.55	0.62	0.56	1210

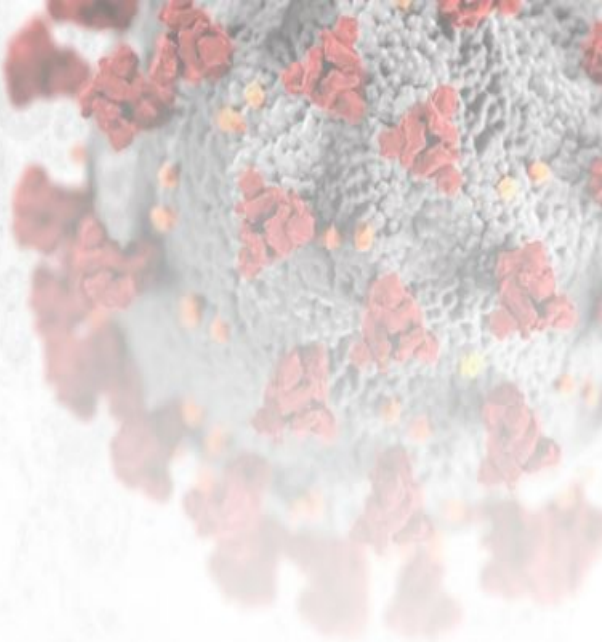
Non Pretrained Resnet50

	precision	recall	f1-score	support
Atypical Appearance	0.00	0.00	0.00	103
Indeterminate Appearance	0.00	0.00	0.00	168
Negative for Pneumonia	0.53	0.61	0.57	353
Typical Appearance	0.60	0.83	0.70	587
accuracy			0.58	1211
macro avg	0.28	0.36	0.32	1211
weighted avg	0.45	0.58	0.50	1211

- **All models cannot detect class ‘Atypical Appearance’ and ‘Indeterminate Appearance’ well. Non pretrained Vgg16 and non pretrained Resnet can not detect any of them**
- **Pretrained Vgg16 model has the highest average f1-score(macro avg), which is 0.40, and its accuracy is 62%**
- **Non pretrained Vgg16 model has the highest accuracy, which is 64%**

Because of class imbalance, f1-score will be the main metric of this project. Therefore, the **Pretrained Vgg16 model** will be the the best one for these image data set.





Conclusion

Conclusion

- **The best model for this chest radiograph set is Vgg16 with the pretrained weights**
- **Vgg16 with the pretrained weights got a 62% accuracy and 0.40 f1-score, which are too far off from the criteria of project success(Accuracy>=90%, f1-score>=0.80)**
- **The reason for failure may be because of the lack of image data or the images in different classes are too similar to each other, but we need further research for these guesses.**
- **For further research, the data provider Foundation for the Promotion of Health and Biomedical Research of Valencia Region(FISABIO) should provide more image data, specially the images in classes 'Atypical Appearance' and 'Indeterminate Appearance'**

The End!!!!

Thank you for watching!!!!