

```
In [133... import pandas as pd
import numpy as np
import os
import matplotlib.pyplot as plt
```

```
In [134... train_label="./dataset/train_label.csv"
test_label="./dataset/test_label.csv"
bbox_list_path="./dataset/BBox_List_2017.csv"
data_entry="./dataset/Data_Entry_2017.csv"

train_meta_data = pd.read_csv(train_label)
test_meta_data = pd.read_csv(test_label)
data_entry= pd.read_csv(data_entry)
bbox_list = pd.read_csv(bbox_list_path)
raw_label = pd.concat([train_meta_data,test_meta_data], ignore_index=True)
raw_label.shape
```

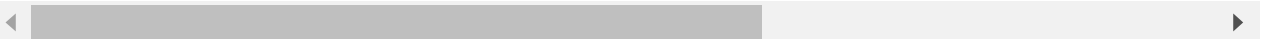
Out[134... (101310, 15)

## 15 classes

```
In [135... label = raw_label.iloc[:,1:]
label.head()
```

```
Out[135... Atelectasis  Cardiomegaly  Effusion  Infiltration  Mass  Nodule  Pneumonia  Pneumothorax  Consolida
```

	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax	Consolidation
0	0	1	0	0	0	0	0	0	
1	0	1	0	0	0	0	0	0	
2	0	1	1	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	
4	0	0	0	0	1	1	0	0	



```
In [136... label.shape
```

Out[136... (101310, 14)

```
In [137... total_No_label_count = 0

for i in range(len(label)):
    if sum(label.iloc[i].values)==0:
        total_No_label_count += 1
```

```
In [138... total_No_label_count
```

Out[138... 53990

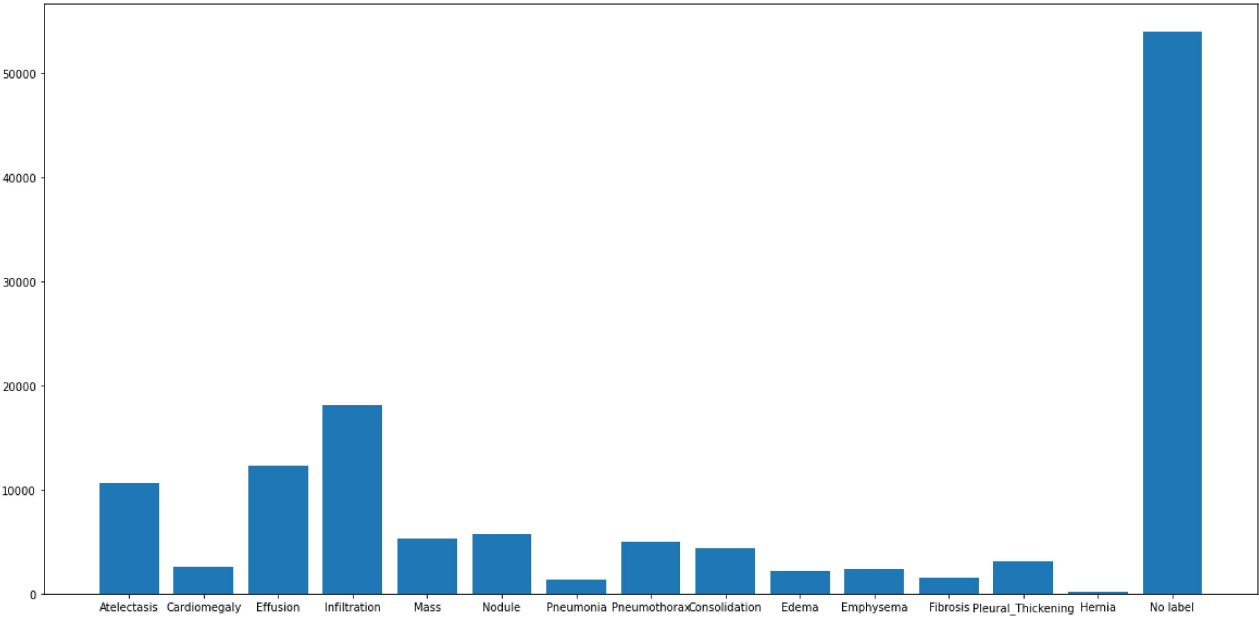
```
In [139... count_of_disease = {}

for name in list(label.columns.values):
    count_of_disease[name] = sum(label[name].values)
```

```
count_of_disease["No label"] = total_No_label_count
count_of_disease
```

```
Out[139... {'Atelectasis': 10585,
'Cardiomegaly': 2559,
'Effusion': 12295,
'Infiltration': 18139,
'Mass': 5327,
'Nodule': 5754,
'Pneumonia': 1317,
'Pneumothorax': 5020,
'Consolidation': 4320,
'Edema': 2143,
'Emphysema': 2399,
'Fibrosis': 1542,
'Pleural_Thickening': 3130,
'Hernia': 212,
'No label': 53990}
```

```
In [140... plt.figure(figsize=(20,10))
plt.bar(*zip(*count_of_disease.items()))
plt.show()
```



8 classes

```
In [141... label = raw_label.iloc[:,1:9]
label.head()
```

```
Out[141...
```

	Atelectasis	Cardiomegaly	Effusion	Infiltration	Mass	Nodule	Pneumonia	Pneumothorax
0	0		1	0	0	0	0	0
1	0		1	0	0	0	0	0
2	0		1	1	0	0	0	0
3	0		0	0	0	0	0	0
4	0		0	0	0	1	1	0

```
In [142... label.shape
```

Out[142... (101310, 8)

```
In [143... total_No_label_count = 0

for i in range(len(label)):
    if sum(label.iloc[i].values)==0:
        total_No_label_count += 1
```

In [144... total\_No\_label\_count

Out[144... 58678

```
In [145... count_of_disease = {}
for name in list(label.columns.values):
    count_of_disease[name] = sum(label[name].values)

count_of_disease["No label"] = total_No_label_count
count_of_disease
```

Out[145... {'Atelectasis': 10585,  
'Cardiomegaly': 2559,  
'Effusion': 12295,  
'Infiltration': 18139,  
'Mass': 5327,  
'Nodule': 5754,  
'Pneumonia': 1317,  
'Pneumothorax': 5020,  
'No label': 58678}

```
In [146... plt.figure(figsize=(20,10))
plt.bar(*zip(*count_of_disease.items()))
plt.show()
```

