CLASSIFICATION OF LUNG INFECTIONS

by: anika rastogi

(NOT ALL)

01

PROBLEM

Lung infections claim numerous lives yearly , causing grief. A classifier can significantly improve early detection and management. A classifier excels in swiftly analyzing medical imaging, like X-rays and CT scans. So my question is:

how can we uses classifier to identify lung disease??? And how do we make it??

METHOD

- works throughout multiple different algorithm
- The process of image recognition

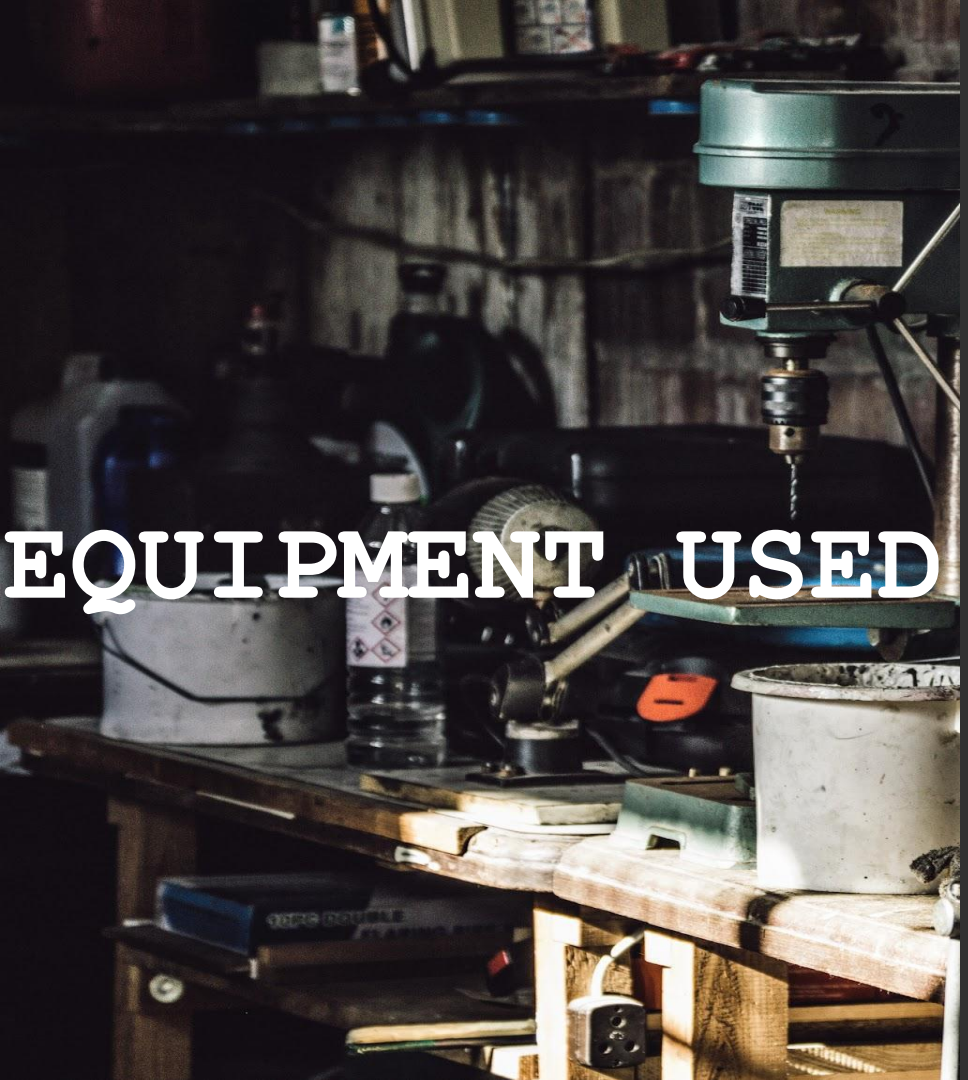
|_> begins with collecting a range of pictures, and label it training data

|_>, we can adjusts the parameters, and feature, so it can minimize the difference between all the photos.

|_> the model is able to recognize any unsee pictures, this is where the testing data comes in.

|_>We can now input a group of photos(test data)

|_>it will make predictions, and give us a percent of accuracy FOR each disease trained for.



EQUIPMENT USED

the equipment used-

- my computer
- a dataset from kaggle,google colab
- python libraries (sklearn,numpy,torch,and matplotlib.pyplot.)
- The algorithm i used was svm

THE DATASET

the dataset i used was worth 2GB

it contained 3 folders - train, test, and val
And those contain, tuberculosis, bacterial pneumonia, viral pneumonia, covid-19, normal files.

basically 10000 images

COMPARING TODAY'S INNOVATIONS

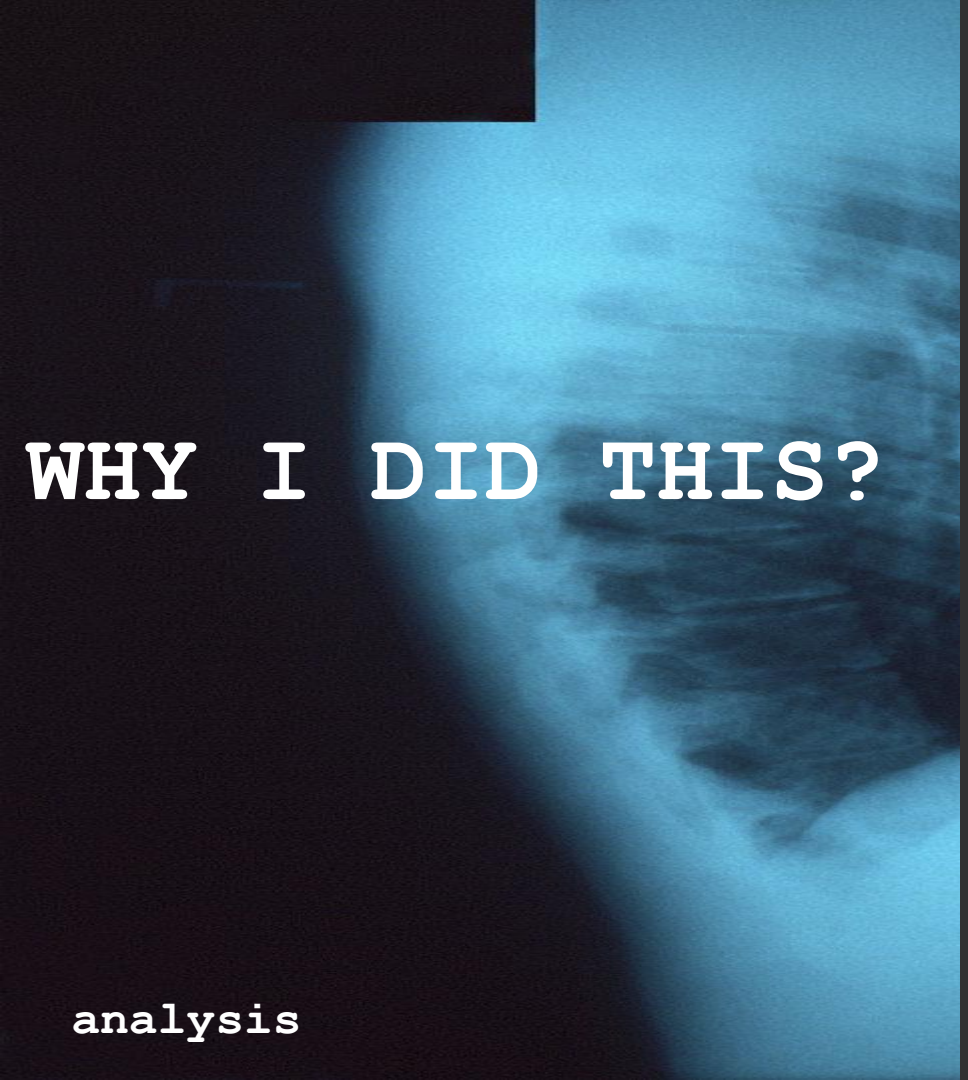
analysis

imaging-

is a diagnostic test, that uses multiple different angles, creating a detailed image of the lung.

Nurses or radiologist or residents look at the imaging, and based on lung doctor, many imaging come back and they seem questionable so it takes many many people to get it right.
|_> PA- when you use a film to view the lung. a beam is shined through your back and it appears through the happy (but only used when patient is really sick
|_> lateral- view the lungs from the side.

There are many people in the medical field who can check lung imaging, but in general(the average) between all is not very good. (based off a real doctor)

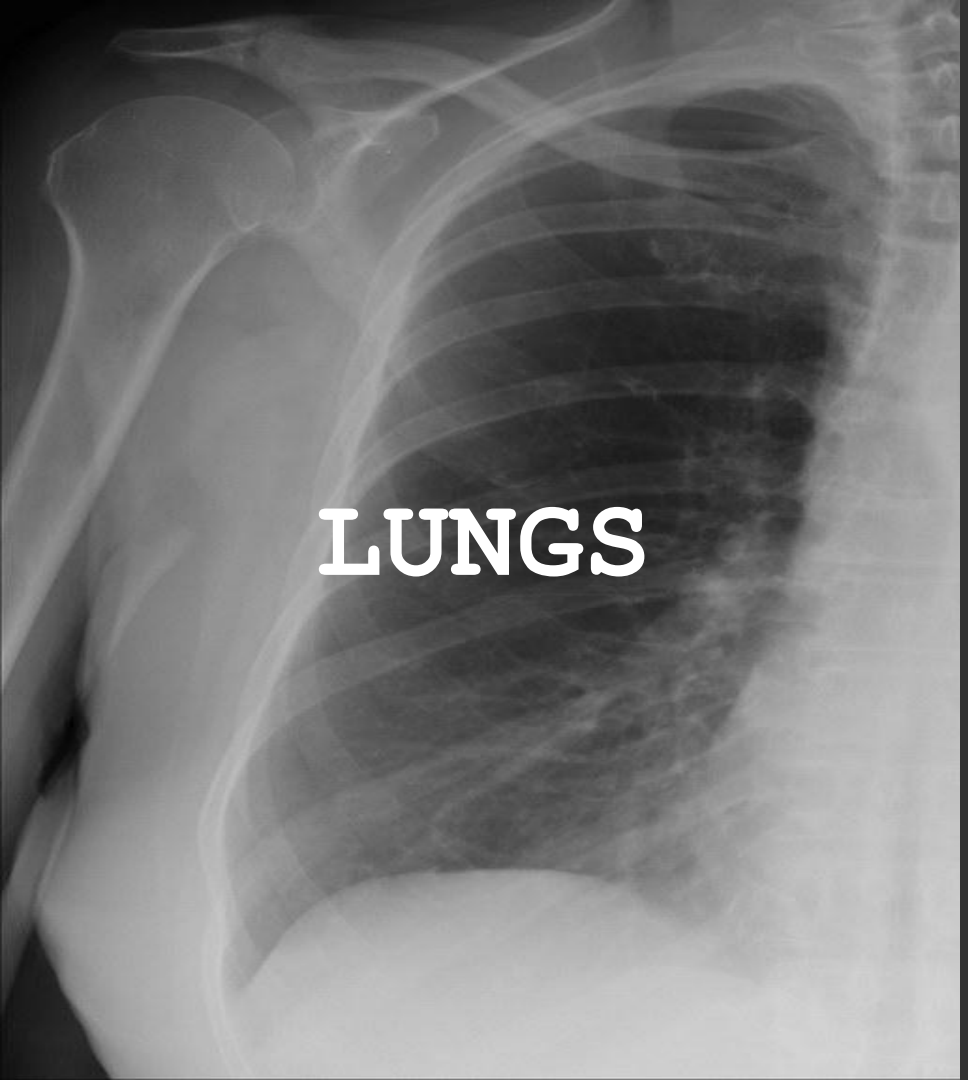


WHY I DID THIS?

analysis

The reason I undertook this project is because many people die from these 4 diseases, viral pneumonia, bacterial pneumonia, tuberculosis, and covid-19.
|_> All of these has caused millions of deaths.
|_>The survival rate is at least 23% with an average of 1.8 million people dying each year.
|_>but sometimes these symptoms are very similar and the only way to identify them are imaging scans,
    |_> this AI (if improved a bit more) can make the job for doctors more efficient and less lives will be lost to lung diseases.

02
research



LUNGS

- Lung are a vital organ in the body(respiratory system).

|_> in exchanging oxygen, and carbon dioxide to sustain life.

|_> millions and millions of sacs called alveoli, which are surrounds by multiple different blood vessels.

- process:

|_> inhalation -> air in arriving toward the lung

|_> exhalation -> the carbon is expelled from the body.

COVID-19

- COVID19, came as a worldwide pandemic, in late 2019.

|_> spreads through the respiratory droplets, causing multiple different symptoms

- .it can be spread in multiple different ways.

|_> its transferred by air

- Many people have faced death due to this disease

TUBERCULOSIS

- Tuberculosis(tb) is a big bacterial infection caused by mycobacterium.

|_> IT mainly affects the lungs, but can 100% affect other organs.

|_> transferred by air

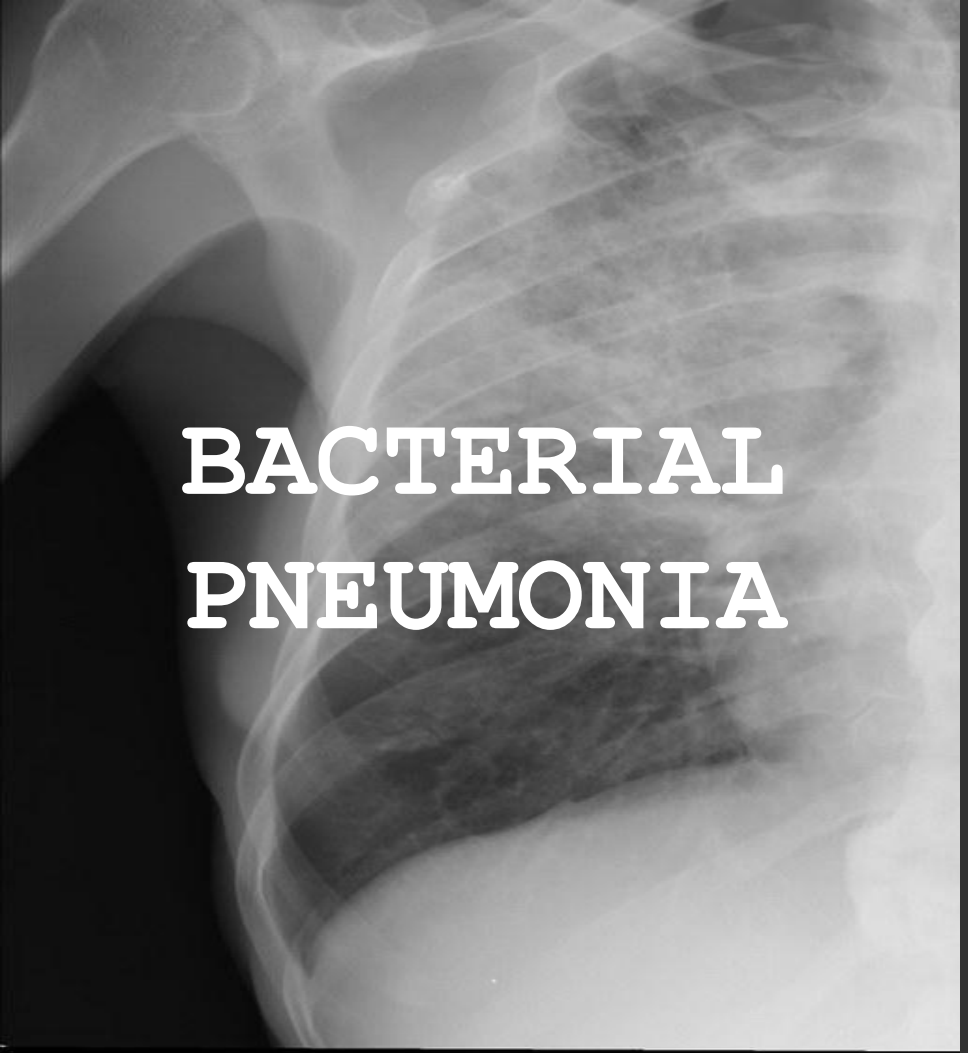
TB has been a longstanding world health concern

|_> testing everyone working in the medical field

|_> still finding a way to efficiently end the disease on the human body.

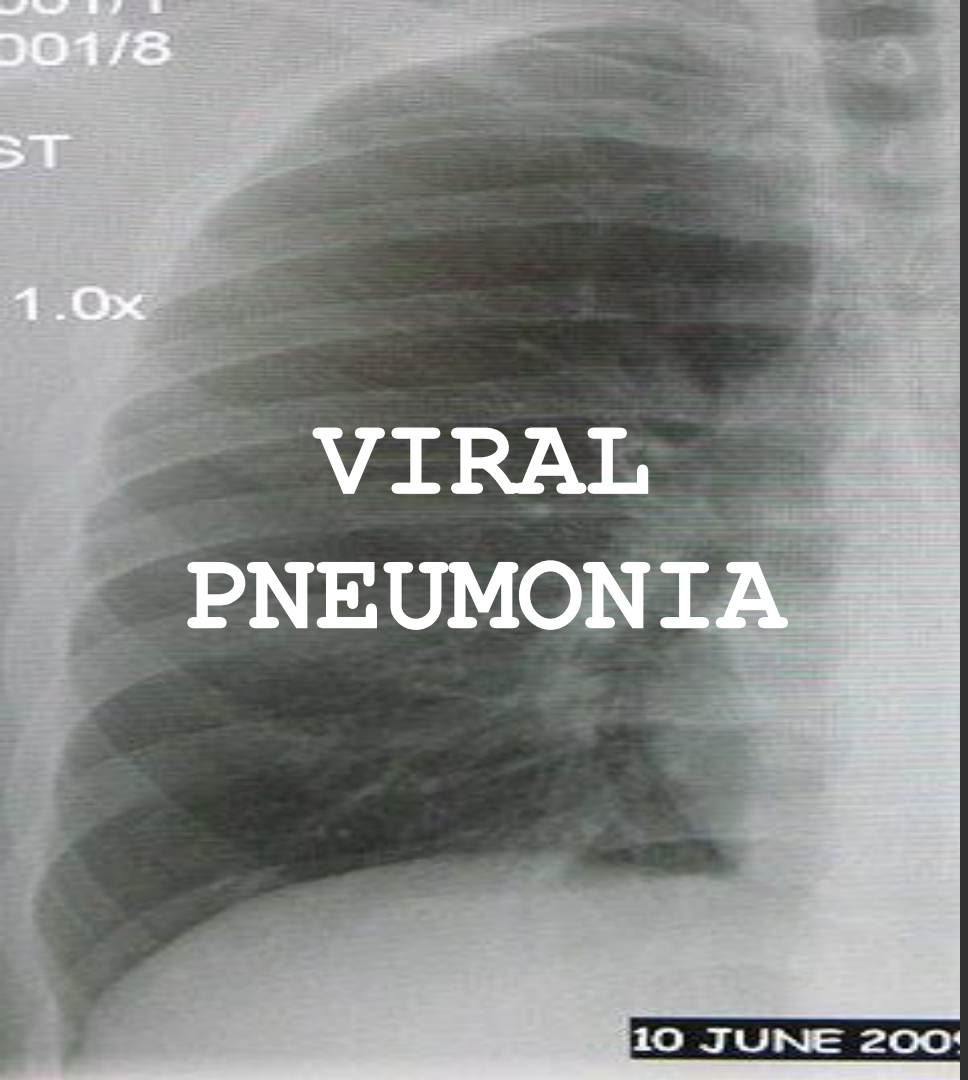
- About a quarter of the world is approximately affected by TB, and this continues.

|_> cured by antibiotics, without antibiotics increased chance of dying.



BACTERIAL PNEUMONIA

- Bacterial pneumonia is a disease
|_> cause inflammation of the lungs, and difficulty breathing
|_> symptoms, like cold, fever, and shortness of breath.
- the main difference between the 2
|_> is BP is more severe and way more common than VP.



VIRAL PNEUMONIA

- Viral pneumonia is the result of infections.

|_> influenza, RSV(respiratory syncytial virus), and even COVID-19.

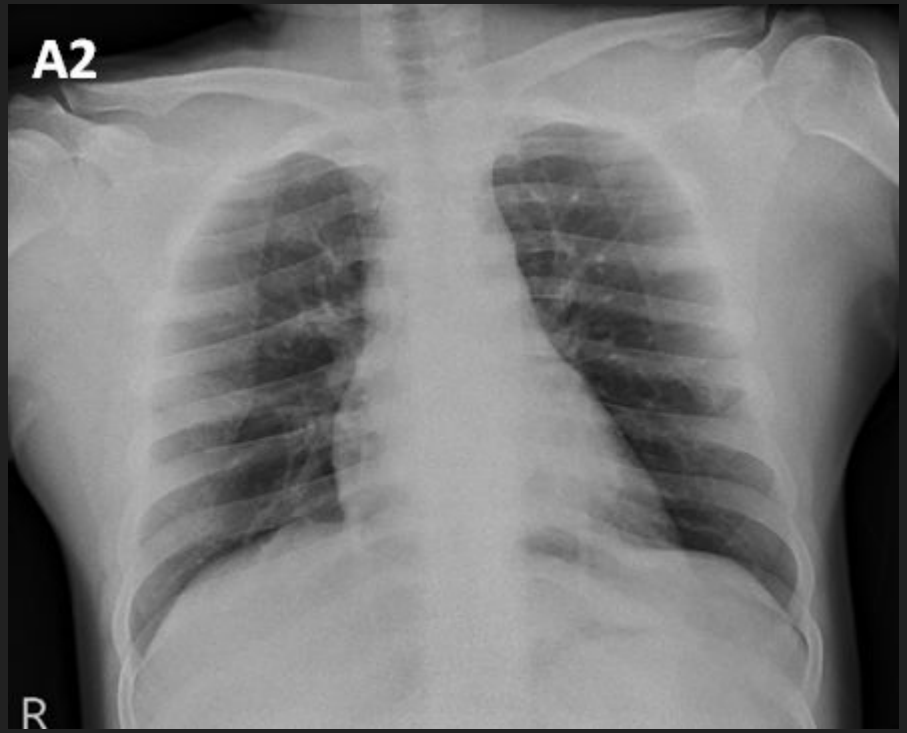
|_> cause inflammation in the lung tissue(making it grow bigger)

- BUT this disease is preventable

|_> by washing your hands, and face(HYGIENE)

|_> taking either the flu, or COVID-19 vaccine

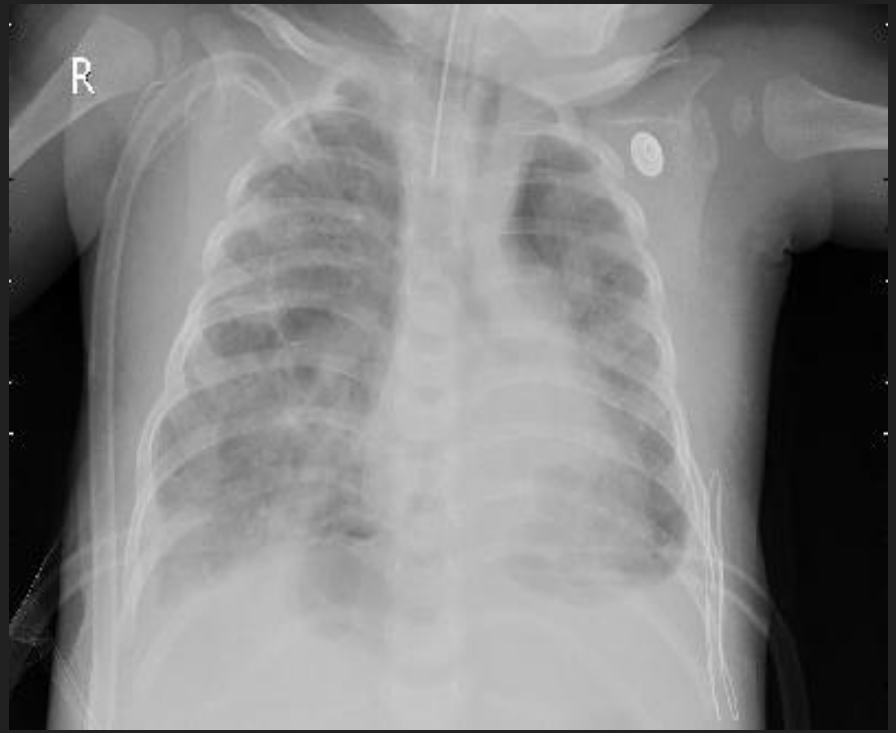
|_> not smoking



COVID-19

TUBERCULOSIS

BACTERIAL PNEUMONIA



VIRAL PNEUMONIA.

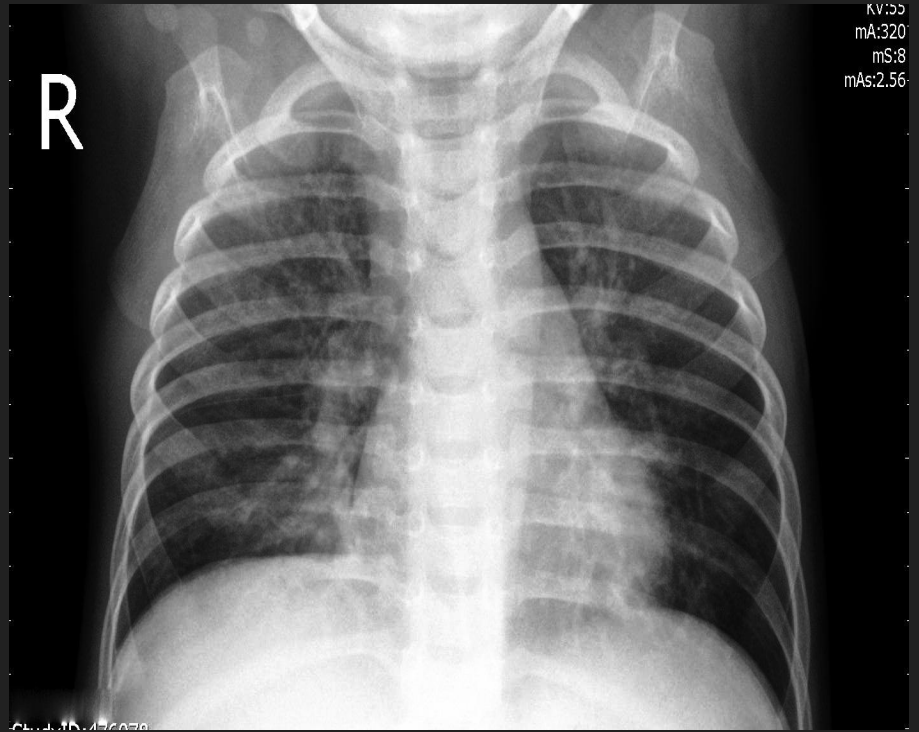
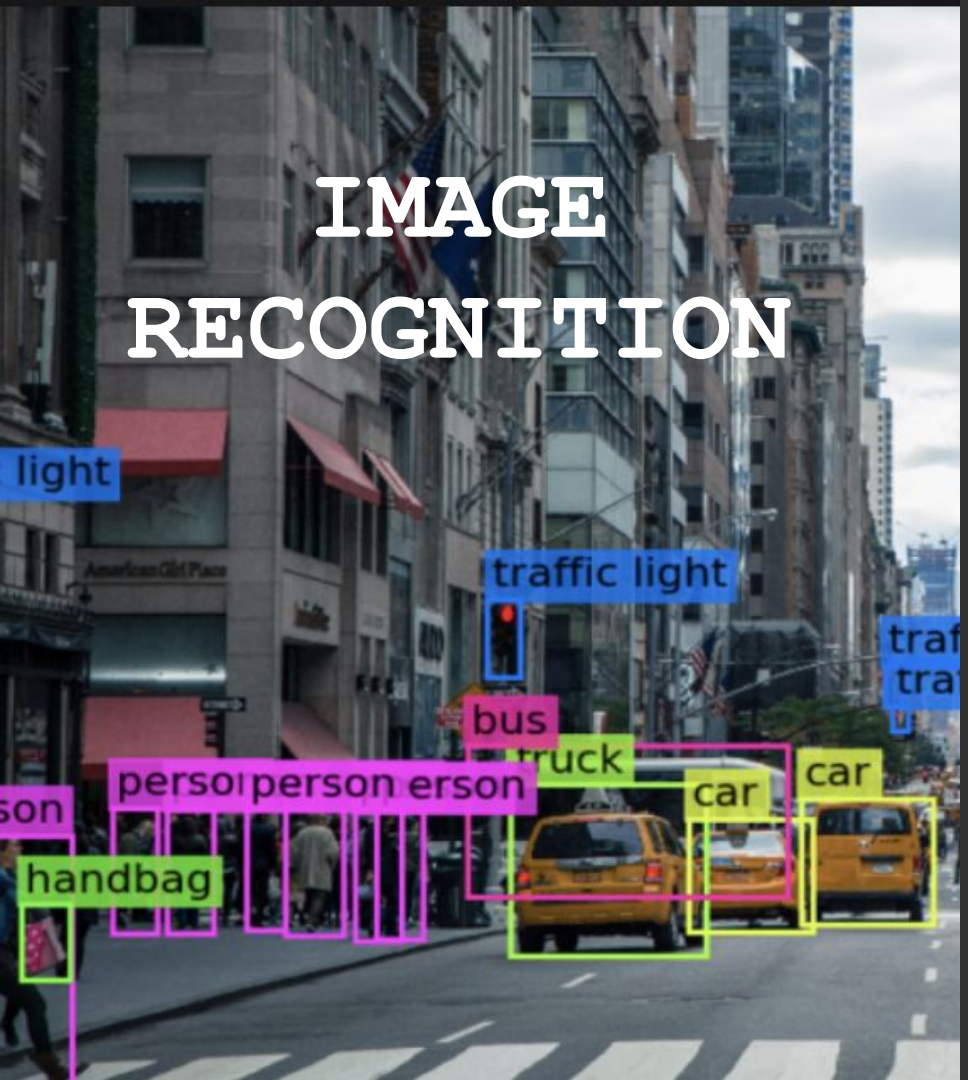


IMAGE RECOGNITION



- Image recognition, aka computer vision,
  |_> gives machines the ability to interpret and recognize visual images.
- to detect patterns and objects in visual designs given to the computer. And the computer/or model create would reduce the difference to see which are the most similar and classify them.

WHAT I DID?

(HOW DOES IT WORK??)

- found a dataset on kaggle consisting of all the data i needed.
- then I uploaded all my data into the code and unzip them
- i import all the libraries(sklearn,numpy, torch, and matplotlib.pyplot.)

- i open the image
|_> convert the image into grayscale
|_> resize(in my casse 300 to 300)
|_> then convert to array
|_> ADD this to the array.
|_> Repeating this process for each disease

example-

```
print("loading normal images")
for filename in normal_img_names[:200]:(-> open the image)
 img = Image.open(f"./train_images/normaltrain/{filename}")
|_> convert to grayscale
 img = ImageOps.grayscale(img)(-> resize to 300x300)
 img = img.resize(size)(-> convert to array)
 imgarr = np.array(img)(-> append to our data array)
 data.append(imgarr)
 target.append("normal")
```

- we create a classification report and then print the report -> creating our precision and accuracy.

RESULTS

```
] print(
      "Classification Report: \n",
      f"{metrics.classification_report(y_test,predicted)}"
  )
```

```
Classification Report:
                     precision    recall  f1-score   support

  baterical pneumonia       0.63      0.76      0.69        55
                covid       0.74      0.75      0.74        64
               normal       0.82      0.71      0.76        58
          tuberculosis       0.69      0.73      0.71        66
       viral pneumonia       0.65      0.54      0.59        57

             accuracy                           0.70       300
            macro avg       0.70      0.70      0.70       300
         weighted avg       0.70      0.70      0.70       300
```

analysis

precision = accuracy of prediction

bacterial pneumonia= *63%*

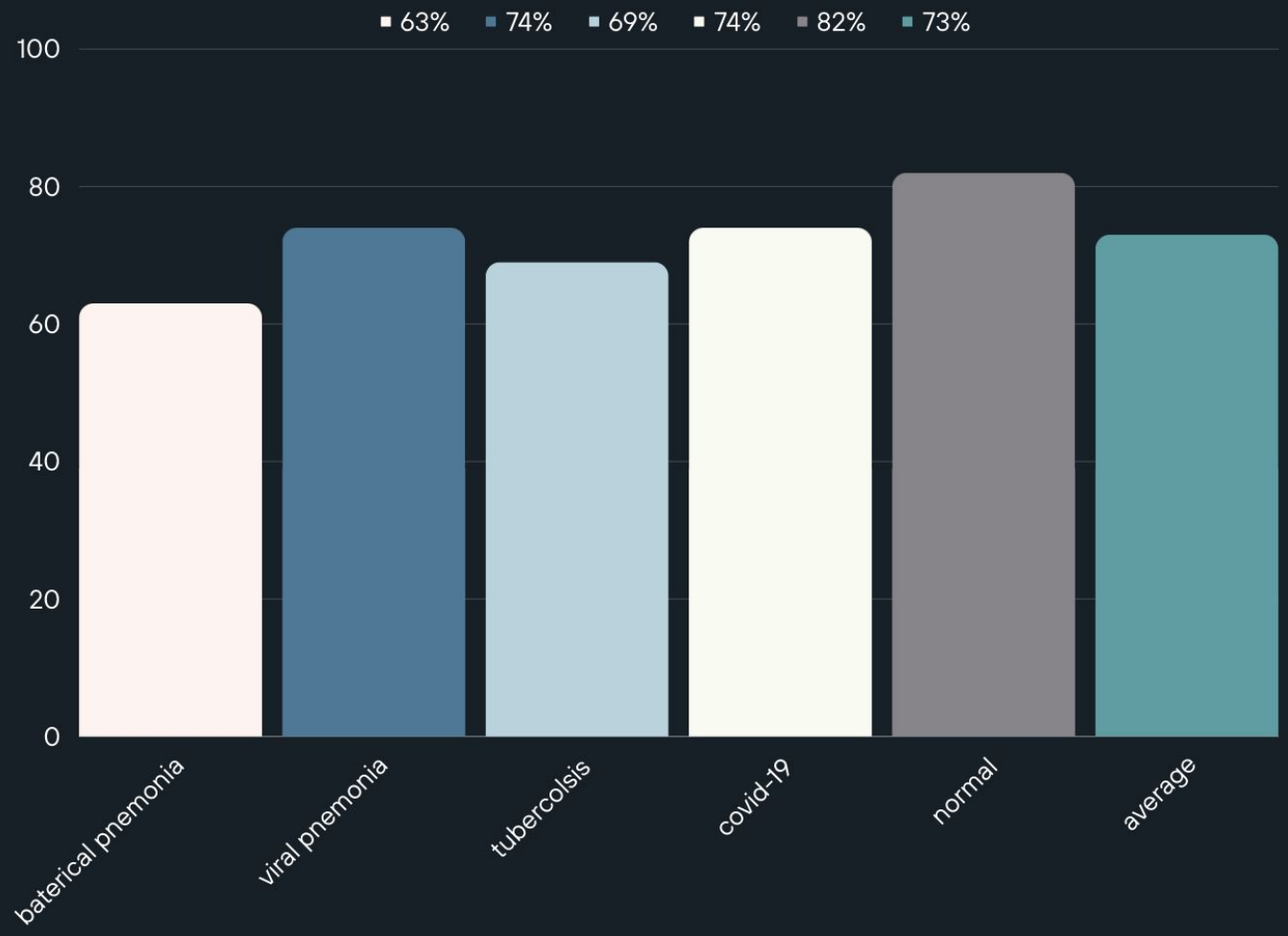
viral pneumonia = *74%*

tuberculosis = *69%*

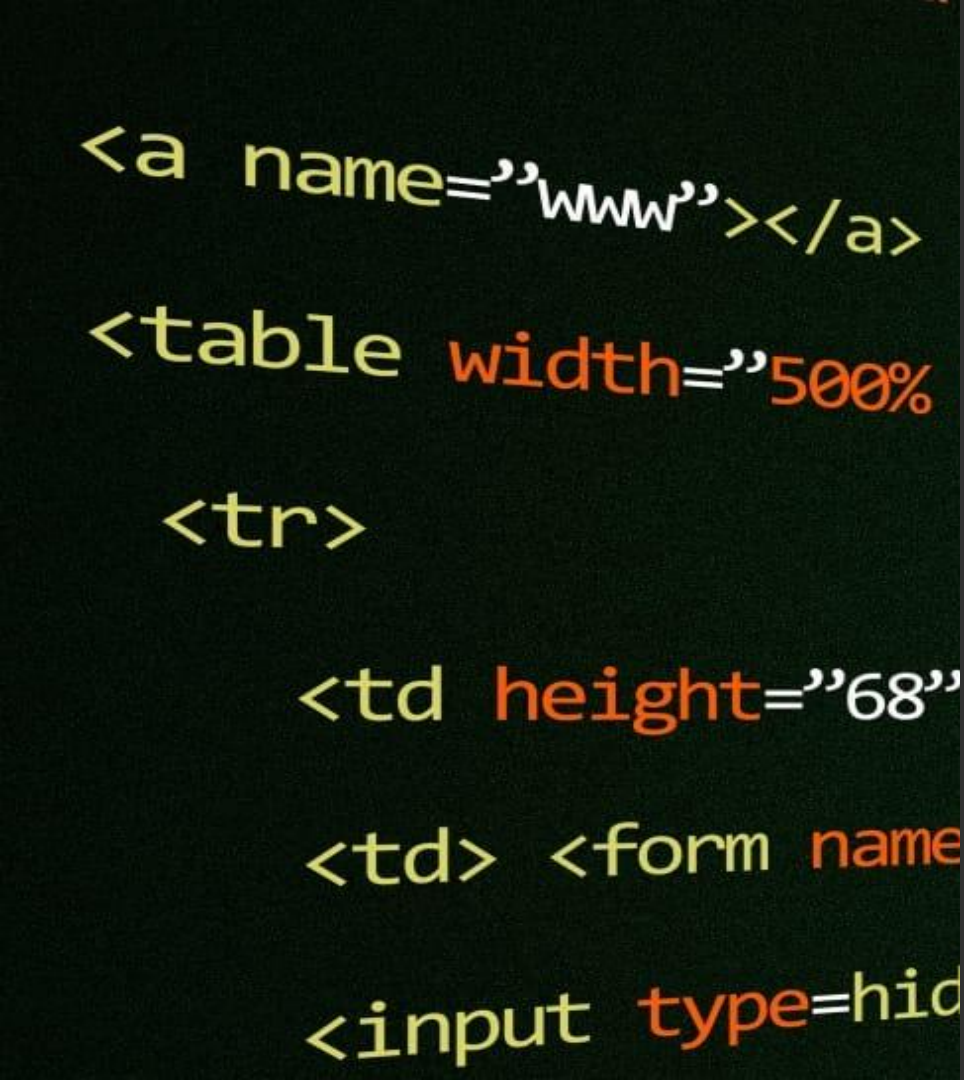
covid-19 = *74%*

normal = *82%*

average = *75%*


63%
74%
69%
74%
82%
73%
100
80
60
40
20
0
baterical pnemonia
viral pnemonia
tubercolsis
covid-19
normal
average

GRAPH



<table width=”500%
<tr>
<td height=”68”
<td> <form name
<input type=hid

03

MY CODE

```
[ ]  !unzip /content/tubetrain.zip
```

```
     from sklearn import metrics,svm
     from sklearn.model_selection import train_test_split

     import matplotlib.pyplot as plt
     from PIL import Image, ImageOps
     import numpy as np
     import os
     import torch
     from torch.utils.data import Dataset
     from torchvision.transforms import ToTensor
```

```
[ ]  normal_img_names = os.listdir("./train_images/normaltrain")
     BP_img_names = os.listdir("./train_images/BPtrain")
     CVD_img_names = os.listdir("./train_images/CVDtrain")
     tuber_img_names = os.listdir("./train_images/tubetrain")
     VP_img_names = os.listdir("./train_images/VPtrain")
```

```
     #img = Image.open("./test_images/BPtest/116.jpeg")
     #img = img.resize((300,300))
     #imgarr = np.array(img)
     #print(imgarr)
     #img
```

```
[ ]  data = []
     target = []

     size = (100,100)

     print("loading normal images")
```

```
     for filename in normal_img_names[:200]:
       # open the image
       img = Image.open(f"./train_images/normaltrain/{filename}")
       # convert to grayscale
       img = ImageOps.grayscale(img)
       # resize to 300x300
       img = img.resize(size)
       # convert to array
       imgarr = np.array(img)
       # append to our data array
       data.append(imgarr)
       target.append("normal")

     print("loading baterical pnemonia Images")
     for filename in BP_img_names[:200]:
       img = Image.open(f"./train_images/BPtrain/{filename}")
       img = ImageOps.grayscale(img)
       img = img.resize(size)
       imgarr = np.array(img)
       data.append(imgarr)
       target.append("baterical pneumonia")

     print("loading covid images")
     for filename in CVD_img_names[:200]:
       img = Image.open(f"./train_images/CVDtrain/{filename}")
       img = ImageOps.grayscale(img)
       img = img.resize(size)
       imgarr = np.array(img)
       data.append(imgarr)
       target.append("covid")

     print("loading viral pnemonia images")
     for filename in VP_img_names[:200]:
       img = Image.open(f"./train_images/VPtrain/{filename}")
       img = ImageOps.grayscale(img)
       img = img.resize(size)
       imgarr = np.array(img)
```

```
    img = img.resize(size)
    imgarr = np.array(img)
    data.append(imgarr)
    target.append("viral pneumonia")

print("loading tuberculosis images")
for filename in tuber_img_names[:200]:
    img = Image.open(f"./train_images/tubetrain/{filename}")
    img = ImageOps.grayscale(img)
    img = img.resize(size)
    imgarr = np.array(img)
    data.append(imgarr)
    target.append("tuberculosis")

data = np.array(data)
target = np.array(target)
num_images = len(data) # get the amount of images in our dataset
data = data.reshape((num_images,-1))
```

```
loading normal images
loading baterical pnemonia Images
loading covid images
loading viral pnemonia images
loading tuberculosis images
```

```
[ ] clf = svm.SVC(verbose=True) # create our classifier

    X_train, X_test, y_train, y_test = train_test_split(data, target, test_size=0.3, shuffle=True) # splitting into a test set and a train set

    clf.fit(X_train, y_train) # this line trains the model on the digits

    predicted = clf.predict(X_test) # makes predictions of the labels on the test set using our new model
```

```
[LibSVM]
```

```
[ ] print(
        "Classification Report: \n",
        f"{metrics.classification_report(y_test,predicted)}"
    )
```

```
Classification Report:
                     precision    recall  f1-score   support

baterical pneumonia       0.63      0.76      0.69        55
              covid       0.74      0.75      0.74        64
             normal       0.82      0.71      0.76        58
       tuberculosis       0.69      0.73      0.71        66
     viral pneumonia       0.65      0.54      0.59        57

           accuracy                           0.70       300
          macro avg       0.70      0.70      0.70       300
       weighted avg       0.70      0.70      0.70       300
```

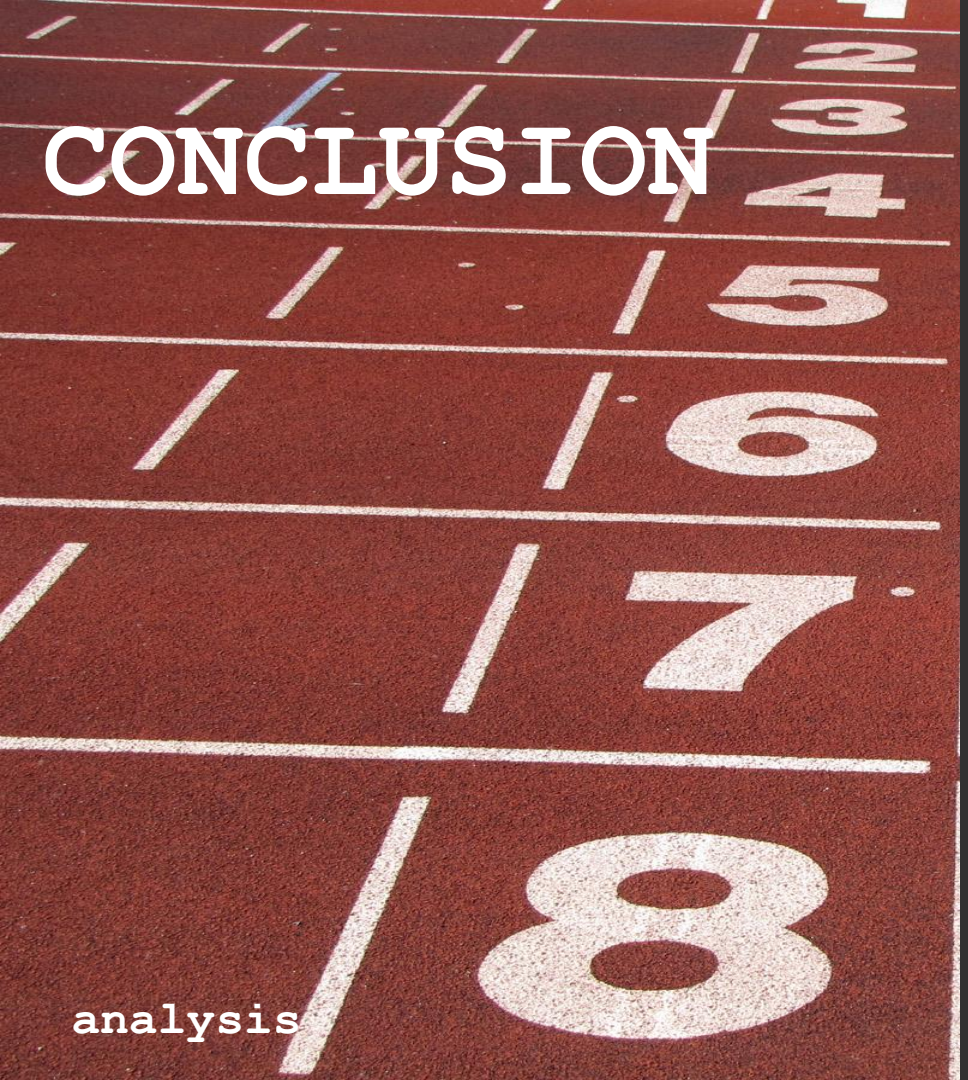
```
print(X_train)
print(y_train)
```

```
[[ 40  39  40 ...  97  78  68]
 [121 123 123 ...  10  11   9]
 [122 104 144 ...   0   0   0]
 ...
 [  0   0   0 ...   5   5   0]
 [  0   0   0 ... 162 163 161]
 [ 35  50  24 ...  55  17  11]]
['baterical pneumonia' 'normal' 'normal' 'viral pneumonia'
 'baterical pneumonia' 'covid' 'tuberculosis' 'covid' 'viral pneumonia'
 'baterical pneumonia' 'covid' 'covid' 'viral pneumonia' 'viral pneumonia'
 'covid' 'covid' 'baterical pneumonia' 'viral pneumonia' 'tuberculosis'
 'viral pneumonia' 'normal' 'tuberculosis' 'normal' 'baterical pneumonia'
 'covid' 'viral pneumonia' 'normal' 'normal' 'normal' 'tuberculosis'
 'baterical pneumonia' 'covid' 'normal' 'tuberculosis' 'viral pneumonia'
 'baterical pneumonia' 'tuberculosis' 'viral pneumonia' 'viral pneumonia'
 'viral pneumonia' 'tuberculosis' 'baterical pneumonia' 'tuberculosis'
 'baterical pneumonia' 'baterical pneumonia' 'tuberculosis' 'tuberculosis'
 'tuberculosis' 'baterical pneumonia' 'covid' 'covid' 'viral pneumonia'
 'covid' 'viral pneumonia' 'tuberculosis' 'baterical pneumonia'
 'viral pneumonia' 'covid' 'viral pneumonia' 'normal' 'viral pneumonia'
 'baterical pneumonia' 'covid' 'normal' 'baterical pneumonia'
 'viral pneumonia' 'normal' 'baterical pneumonia' 'covid'
 'baterical pneumonia' 'baterical pneumonia' 'viral pneumonia'
 'viral pneumonia' 'normal' 'normal' 'tuberculosis' 'normal'
 'tuberculosis' 'normal' 'tuberculosis' 'viral pneumonia'
 'baterical pneumonia' 'baterical pneumonia' 'normal' 'covid' 'covid'
 'tuberculosis' 'normal' 'viral pneumonia' 'normal' 'normal' 'covid'
 'baterical pneumonia' 'covid' 'tuberculosis' 'covid' 'normal'
 'baterical pneumonia' 'normal' 'covid']
```

```
print(X_test)
```

```
[[ 62  82  98 ...  11  19  21]
 [117 115 131 ...  23   0   0]
 [ 57  73  84 ...  32  33  33]
 ...
 [  6   6   6 ... 178 170 166]
 [ 16  17  17 ...  16  19  19]
 [  0   0  11 ...   0   0   0]]
```

```
[ ] plt.imshow(train_images[0], cmap='gray')
    plt.show()
```



CONCLUSION

- SUMMARY
- what could happen if improved-
  |_> hospitals
    |_> further in the world.
- if improved what could happen
  |_> death rates

FUTURE IMPROVEMENT

- adding more algorithms
  |_> RCNN, …
- gather bigger datasets
  |_> mix a few together.
- add more lung infections, when improved further.
- Instead of only putting one angle- put multiple different angles - like lateral.
  |_> lung cancer
    |_> like adenocarcinoma, squamous cell carcinoma, and many more.

REAL-LIFE APPLICATION

Used to check work of residents: residents are the back bone of the hospital because they work the night shifts, and sometime xrays come in the middle of the night,and they are not fully prepared to make that official decision. So an ai can assist them.

can improve giving unnecessary antibiotics -> many are given from false diagoness.

radiologist-> unnecessary money.

low resource country-> improved more it could become a work horse, and improve the world globally.

Pulling normal images- pulling normals ones out from diseased (narrow search)

CITATION

AI Plain English. "Creating a Neural Network for Image Recognition with Keras in Python." Medium, Plain English,https://ai.plainenglish.io/creating-a-neural-network-for-image-recognition-with-keras-in-python-1f285bc48f13 .

V7 Labs. "Image Recognition Guide." V7 Labs,https://www.v7labs.com/blog/image-recognition-guide .

TIBCO Spotfire. "What is a Neural Network?" Spotfire, https://www.spotfire.com/glossary/what-is-a-neural-network .

TechTarget. "Image Recognition." TechTarget,https://www.techtarget.com/searchenterpriseai/definition/image-recognition#:~:text=For%20example%2C%20when%20someone%20uploads,the%20individual%20in%20the%20image

Mayo Clinic. "Tuberculosis - Symptoms & causes." Mayo Clinic, https://www.mayoclinic.org/diseases-conditions/tuberculosis/symptoms-causes/syc-20351250 .

American Lung Association. "Pneumonia Symptoms and Diagnosis." American Lung Association, https://www.lung.org/lung-health-diseases/lung-disease-lookup/pneumonia/symptoms-and-diagnosis .

WebMD. "Bacterial Pneumonia: Symptoms, Causes, Treatment." WebMD, https://www.webmd.com/lung/bacterial-pneumonia.

Centers for Disease Control and Prevention. "Symptoms of COVID-19." CDC, https://www.cdc.gov/coronavirus/2019-ncov/symptoms-testing/symptoms.html.

Kaggle. "Lungs Disease Dataset (4 Types)." Kaggle, https://www.kaggle.com/datasets/omkarmanohardalvi/lungs-disease-dataset-4-types .

ACKNOWLEGDEMENT

I acknowledge my mother, my father, and my dearest awesome brother.

I acknowledge the people from juniotech, who helped me along the way, especially irada, and tim

I acknowledge ms.lai for giving me this opportunity

And lastly i acknowledge Izabela Sztukowski for helping me with my research on lungs.

thank you!!