In [0]:

```
!pip install kaggle
```

```
Requirement already satisfied: kaggle in /usr/local/lib/python3.6/dist-packages (1.5.6)
Requirement already satisfied: python-slugify in /usr/local/lib/python3.6/dist-packages (from kaggle
) (4.0.0)
Requirement already satisfied: python-dateutil in /usr/local/lib/python3.6/dist-packages (from kaggl
e) (2.8.1)
Requirement already satisfied: six>=1.10 in /usr/local/lib/python3.6/dist-packages (from kaggle) (1.
12.0)
Requirement already satisfied: certifi in /usr/local/lib/python3.6/dist-packages (from kaggle) (2020
.4.5.1)
Requirement already satisfied: urllib3<1.25,>=1.21.1 in /usr/local/lib/python3.6/dist-packages (from
kaggle) (1.24.3)
Requirement already satisfied: tqdm in /usr/local/lib/python3.6/dist-packages (from kaggle) (4.41.1)
Requirement already satisfied: requests in /usr/local/lib/python3.6/dist-packages (from kaggle) (2.2
3.0)
Requirement already satisfied: text-unidecode>=1.3 in /usr/local/lib/python3.6/dist-packages (from p
ython-slugify->kaggle) (1.3)
Requirement already satisfied: idna<3,>=2.5 in /usr/local/lib/python3.6/dist-packages (from requests
->kaggle) (2.9)
Requirement already satisfied: chardet<4,>=3.0.2 in /usr/local/lib/python3.6/dist-packages (from req
uests->kaggle) (3.0.4)
```

In [0]:

```
# Upload your kaggel api file here
# For reference watch these video link :- https://www.youtube.com/watch?v=tGw-ZACouik
from google.colab import files
files.upload()
```

In [0]:

```
!mkdir -p ~/.kaggle
!cp kaggle.json ~/.kaggle
!chmod 600 ~/.kaggle/kaggle.json
```

In [0]:

```
# import the necessary packages
from tensorflow.keras.preprocessing.image import ImageDataGenerator
from tensorflow.keras.applications import VGG16
from tensorflow.keras.layers import AveragePooling2D
from tensorflow.keras.layers import Dropout
from tensorflow.keras.layers import Flatten
from tensorflow.keras.layers import Dense
from tensorflow.keras.layers import Input
from tensorflow.keras.models import Model
from tensorflow.keras.optimizers import Adam
from tensorflow.keras.utils import to_categorical
from sklearn.preprocessing import LabelBinarizer
from sklearn.model_selection import train_test_split
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from imutils import paths
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
import random
import shutil
import cv2
import os
```

In [0]:

```
!kaggle datasets download -d paultimothymooney/chest-xray-pneumonia
```

```
Downloading chest-xray-pneumonia.zip to /content
100% 2.28G/2.29G [00:34<00:00, 50.8MB/s]
100% 2.29G/2.29G [00:34<00:00, 72.1MB/s]
```

In [0]:

```
!ls
```

```
chest-xray-pneumonia.zip  kaggle.json  sample_data
```

In [0]:

```
from zipfile import ZipFile
filename = "chest-xray-pneumonia.zip"

with ZipFile(filename,'r') as zip:
  zip.extractall()
  print('Done')
```

```
Done
```

In [0]:

```
!kaggle datasets download -d bachrr/covid-chest-xray
```

```
Downloading covid-chest-xray.zip to /content
100% 240M/241M [00:05<00:00, 44.8MB/s]
100% 241M/241M [00:05<00:00, 44.5MB/s]
```

In [0]:

```
from zipfile import ZipFile
filename = "covid-chest-xray.zip"

with ZipFile(filename,'r') as zip:
  zip.extractall()
  print('Done')
```

```
Done
```

In [0]:

```
!mkdir -p ~/.dataset
```

In [0]:

```
!ls
```

```
annotations  chest-xray-pneumonia.zip  images       metadata.csv
chest_xray   covid-chest-xray.zip      kaggle.json  sample_data
```

In [0]:

```
dataset_path = './dataset'
```

In [0]:

```
%%bash
rm -rf dataset
mkdir -p dataset/covid
mkdir -p dataset/normal
```

In [0]:

```
!ls
```

```
annotations  chest-xray-pneumonia.zip  dataset  kaggle.json   sample_data
chest_xray   covid-chest-xray.zip      images   metadata.csv
```

In [0]:

```
samples = 25
```

In [0]:

```
covid_dataset_path = '/content'
```

In [0]:

```
# construct the path to the metadata CSV file and load it
csvPath = os.path.sep.join([covid_dataset_path, "metadata.csv"])

df = pd.read_csv(csvPath)

# loop over the rows of the COVID-19 data frame
for (i, row) in df.iterrows():
    # if (1) the current case is not COVID-19 or (2) this is not
    # a 'PA' view, then ignore the row
    if row["finding"] != "COVID-19" or row["view"] != "PA":
        continue

    # build the path to the input image file
    imagePath = os.path.sep.join([covid_dataset_path, "images", row["filename"]])

    # if the input image file does not exist (there are some errors in
    # the COVID-19 metadeta file), ignore the row
    if not os.path.exists(imagePath):
        continue

    # extract the filename from the image path and then construct the
    # path to the copied image file
    filename = row["filename"].split(os.path.sep)[-1]
    outputPath = os.path.sep.join([f"{dataset_path}/covid", filename])

    # copy the image
    shutil.copy2(imagePath, outputPath)
```

Build normal xray dataset

In [0]:

```
pneumonia_dataset_path ='/content/chest_xray'
```

In [0]:

```
basePath = os.path.sep.join([pneumonia_dataset_path, "train", "NORMAL"])
imagePaths = list(paths.list_images(basePath))

# randomly sample the image paths
random.seed(42)
random.shuffle(imagePaths)
imagePaths = imagePaths[:samples]

# loop over the image paths
for (i, imagePath) in enumerate(imagePaths):
    # extract the filename from the image path and then construct the
    # path to the copied image file
    filename = imagePath.split(os.path.sep)[-1]
    outputPath = os.path.sep.join([f"{dataset_path}/normal", filename])

    # copy the image
    shutil.copy2(imagePath, outputPath)
```

Plot x-rays

In [0]:

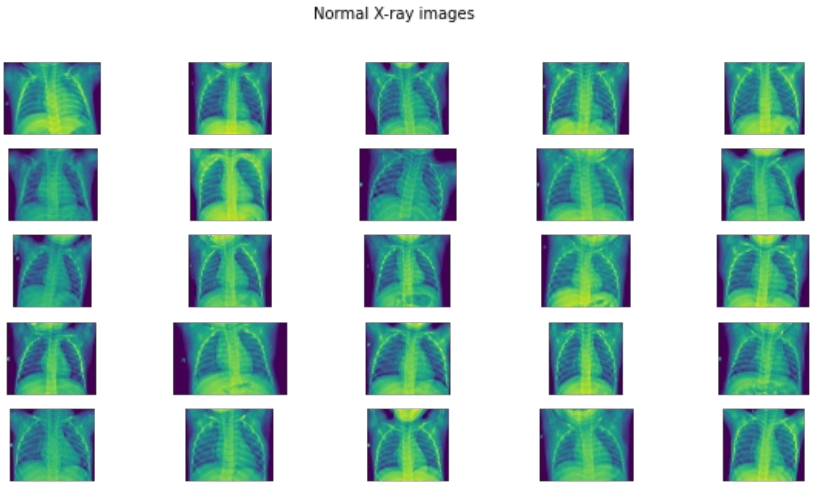
```
def ceildiv(a, b):
    return -(-a // b)

def plots_from_files(imspaths, figsize=(10,5), rows=1, titles=None, maintitle=None):
    """Plot the images in a grid"""
    f = plt.figure(figsize=figsize)
    if maintitle is not None: plt.suptitle(maintitle, fontsize=10)
    for i in range(len(imspaths)):
        sp = f.add_subplot(rows, ceildiv(len(imspaths), rows), i+1)
        sp.axis('Off')
        if titles is not None: sp.set_title(titles[i], fontsize=16)
        img = plt.imread(imspaths[i])
        plt.imshow(img)
```

In [0]:

```
normal_images = list(paths.list_images(f"{dataset_path}/normal"))
covid_images = list(paths.list_images(f"{dataset_path}/covid"))
```

In [0]:

```
plots_from_files(normal_images, rows=5, maintitle="Normal X-ray images")
```




In [0]:

```
plots_from_files(covid_images, rows=5, maintitle="Covid-19 X-ray images")
```


Data preprocessing

In [0]:

```
# initialize the initial learning rate, number of epochs to train for,
# and batch size
INIT_LR = 1e-3
EPOCHS = 10
BS = 8
```

In [0]:

```
# grab the list of images in our dataset directory, then initialize
# the list of data (i.e., images) and class images
print("[INFO] loading images...")
imagePaths = list(paths.list_images(dataset_path))
data = []
labels = []
# loop over the image paths
for imagePath in imagePaths:
    # extract the class label from the filename
    label = imagePath.split(os.path.sep)[-2]
    # load the image, swap color channels, and resize it to be a fixed
    # 224x224 pixels while ignoring aspect ratio
    image = cv2.imread(imagePath)
    image = cv2.cvtColor(image, cv2.COLOR_BGR2RGB)
    image = cv2.resize(image, (224, 224))
    # update the data and labels lists, respectively
    data.append(image)
    labels.append(label)
# convert the data and labels to NumPy arrays while scaling the pixel
# intensities to the range [0, 1]
data = np.array(data) / 255.0
labels = np.array(labels)
```

```
[INFO] loading images...
```

In [0]:

```
# perform one-hot encoding on the labels
lb = LabelBinarizer()
labels = lb.fit_transform(labels)
labels = to_categorical(labels)
# partition the data into training and testing splits using 80% of
# the data for training and the remaining 20% for testing
(trainX, testX, trainY, testY) = train_test_split(data, labels, test_size=0.20, stratify=labels, random_state=42)
# initialize the training data augmentation object
trainAug = ImageDataGenerator(rotation_range=15, fill_mode="nearest")
```

Model

In [0]:

```
baseModel = VGG16(weights="imagenet", include_top=False, input_tensor=Input(shape=(224, 224, 3)))
# construct the head of the model that will be placed on top of the
# the base model
headModel = baseModel.output
headModel = AveragePooling2D(pool_size=(4, 4))(headModel)
headModel = Flatten(name="flatten")(headModel)
headModel = Dense(64, activation="relu")(headModel)
headModel = Dropout(0.5)(headModel)
headModel = Dense(2, activation="softmax")(headModel)
# place the head FC model on top of the base model (this will become
# the actual model we will train)
model = Model(inputs=baseModel.input, outputs=headModel)
# loop over all layers in the base model and freeze them so they will
# *not* be updated during the first training process
for layer in baseModel.layers:
    layer.trainable = False
```

Training

In [0]:

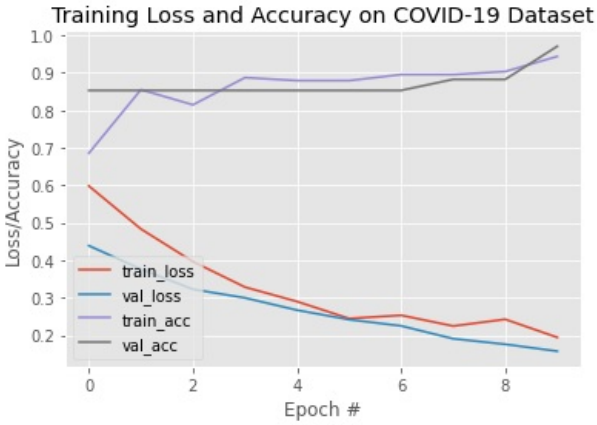
```
# compile our model
print("[INFO] compiling model...")
opt = Adam(lr=INIT_LR, decay=INIT_LR / EPOCHS)
model.compile(loss="binary_crossentropy", optimizer=opt, metrics=["accuracy"])

# train the head of the network
print("[INFO] training head...")
H = model.fit_generator(
    trainAug.flow(trainX, trainY, batch_size=BS),
    steps_per_epoch=len(trainX) // BS,
    validation_data=(testX, testY),
    validation_steps=len(testX) // BS,
    epochs=EPOCHS)
```

```
[INFO] compiling model...
[INFO] training head...
WARNING:tensorflow:From <ipython-input-81-a8d08b9c183d>:13: Model.fit_generator (from tensorflow.pyt
hon.keras.engine.training) is deprecated and will be removed in a future version.
Instructions for updating:
Please use Model.fit, which supports generators.
Epoch 1/10
16/16 [==============================] - 3s 167ms/step - loss: 0.5981 - accuracy: 0.6855 - val_loss:
0.4389 - val_accuracy: 0.8529
Epoch 2/10
16/16 [==============================] - 2s 99ms/step - loss: 0.4834 - accuracy: 0.8548 - val_loss:
0.3764 - val_accuracy: 0.8529
Epoch 3/10
16/16 [==============================] - 2s 104ms/step - loss: 0.3964 - accuracy: 0.8145 - val_loss:
0.3226 - val_accuracy: 0.8529
Epoch 4/10
16/16 [==============================] - 2s 100ms/step - loss: 0.3283 - accuracy: 0.8871 - val_loss:
0.2997 - val_accuracy: 0.8529
Epoch 5/10
16/16 [==============================] - 2s 101ms/step - loss: 0.2900 - accuracy: 0.8790 - val_loss:
0.2670 - val_accuracy: 0.8529
Epoch 6/10
16/16 [==============================] - 2s 100ms/step - loss: 0.2447 - accuracy: 0.8790 - val_loss:
0.2418 - val_accuracy: 0.8529
Epoch 7/10
16/16 [==============================] - 2s 99ms/step - loss: 0.2533 - accuracy: 0.8952 - val_loss:
0.2250 - val_accuracy: 0.8529
Epoch 8/10
16/16 [==============================] - 2s 100ms/step - loss: 0.2246 - accuracy: 0.8952 - val_loss:
0.1908 - val_accuracy: 0.8824
Epoch 9/10
16/16 [==============================] - 2s 99ms/step - loss: 0.2425 - accuracy: 0.9032 - val_loss:
0.1760 - val_accuracy: 0.8824
Epoch 10/10
16/16 [==============================] - 2s 97ms/step - loss: 0.1944 - accuracy: 0.9435 - val_loss:
0.1575 - val_accuracy: 0.9706
```

Plot trining metrics

In [0]:

```
# plot the training loss and accuracy
N = EPOCHS
plt.style.use("ggplot")
plt.figure()
plt.plot(np.arange(0, N), H.history["loss"], label="train_loss")
plt.plot(np.arange(0, N), H.history["val_loss"], label="val_loss")
plt.plot(np.arange(0, N), H.history["accuracy"], label="train_acc")
plt.plot(np.arange(0, N), H.history["val_accuracy"], label="val_acc")
plt.title("Training Loss and Accuracy on COVID-19 Dataset")
plt.xlabel("Epoch #")
plt.ylabel("Loss/Accuracy")
plt.legend(loc="lower left")
plt.savefig("plot.png")
```




Evaluation

In [0]:

```
# make predictions on the testing set
print("[INFO] evaluating network...")
predIdxs = model.predict(testX, batch_size=BS)
# for each image in the testing set we need to find the index of the
# label with corresponding largest predicted probability
predIdxs = np.argmax(predIdxs, axis=1)
# show a nicely formatted classification report
print(classification_report(testY.argmax(axis=1), predIdxs, target_names=lb.classes_))
```

```
[INFO] evaluating network...
              precision    recall  f1-score   support

       covid       0.97      1.00      0.98        29
      normal       1.00      0.80      0.89         5

    accuracy                           0.97        34
   macro avg       0.98      0.90      0.94        34
weighted avg       0.97      0.97      0.97        34
```

Confusion matrix

In [0]:

```
# compute the confusion matrix and and use it to derive the raw
# accuracy, sensitivity, and specificity
cm = confusion_matrix(testY.argmax(axis=1), predIdxs)
total = sum(sum(cm))
acc = (cm[0, 0] + cm[1, 1]) / total
sensitivity = cm[0, 0] / (cm[0, 0] + cm[0, 1])
specificity = cm[1, 1] / (cm[1, 0] + cm[1, 1])
# show the confusion matrix, accuracy, sensitivity, and specificity
print(cm)
print("acc: {:.4f}".format(acc))
print("sensitivity: {:.4f}".format(sensitivity))
print("specificity: {:.4f}".format(specificity))
```

```
[[29  0]
 [ 1  4]]
acc: 0.9706
sensitivity: 1.0000
specificity: 0.8000
```