

DATA SCIENCE CAPSTONE PROJECT EXPLANATION

SARS-COV-2 Ct-Scan Dataset

A large dataset of CT scans for SARS-CoV-2 (COVID-19) identification

Data Overview

This dataset contains 1252 CT scans that are positive for SARS-CoV-2 infection (COVID-19) and 1230 CT scans for patients non-infected by SARS-CoV-2, 2482 CT scans in total. These data have been collected from real patients in hospitals from Sao Paulo, Brazil. The aim of this dataset is to encourage the research and development of artificial intelligent methods which are able to identify if a person is infected by SARS-CoV-2 through the analysis of his/her CT scans.

WORKING ON PROJECT

The image displays a Google Drive interface and a Jupyter Notebook. The Google Drive interface shows a folder named 'Data science Project' containing several Jupyter Notebook files. The Jupyter Notebook shows code for importing libraries, mounting the drive, and loading a CT scan image.

Google Drive Interface:

| Name | Owner | Last modified | File size |
|---|-------|-----------------|-----------|
| Data science Project | me | 1:41 AM me | — |
| Capston covid and non covid project (Ankur Bisen).ipynb | me | 5:32 AM me | 8 KB |
| Copy of Welcome To Colaboratory | me | 1:49 AM me | 125 KB |
| notebook5e3a2127f3.ipynb | me | 5:36 AM me | 16 KB |
| Satyam Tripathi.ipynb | me | 3:25 AM me | 52 KB |
| Untitled | me | Dec 20, 2022 me | 306 bytes |
| Untitled0.ipynb | me | Dec 20, 2022 me | 324 bytes |
| Untitled1.ipynb | me | Dec 20, 2022 me | 429 bytes |
| Untitled2.ipynb | me | Feb 7, 2023 me | 223 bytes |
| Untitled3.ipynb | me | Feb 11, 2023 me | 324 bytes |
| Untitled4.ipynb | me | Feb 11, 2023 me | 662 bytes |

Jupyter Notebook Interface:

```
[18] from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

[19] import numpy as np
import pandas as pd
from pathlib import Path
import os.path
import os
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.model_selection import train_test_split
import glob

[21] '/content/drive/MyDrive/Colab Notebooks/Data science Project/COVID/Covid (1).png'
'/content/drive/MyDrive/Colab Notebooks/Data science Project/COVID/Covid (1).png'

[22] df = pd.DataFrame({'File_path': ['/content/drive/MyDrive/Colab Notebooks/Data science Project/COVID/Covid (1).png']})

df
File_path
0 /content/drive/MyDrive/Colab Notebooks/Data science Project/COVID/Covid (1).png
```

Google
Untitled9.ipynb - Colaboratory
Data science Project - Google Drive

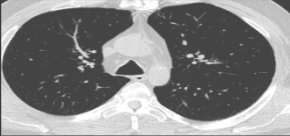
https://colab.research.google.com/drive/1QweoX79kHwSp56_h0QOG2SLvNMN18/#scrollTo=ZEq5C9_dNYE

Import favorites | ASUS Software Port... | MyASUS Software... | McAfee LiveSafe | Can't sign up my d... | RPCS3 - Compatibil... | New tabs

File Edit View Insert Runtime Tools Help All changes saved

+ Code + Text

COVID-19 CT scan images have been collected from hospitals in Singapore from early 2020, where the aim of this dataset is to encourage the research and development of artificial intelligent methods which are able to identify if a person is infected by SARS-CoV-2 through the analysis of his/her CT scans.



upload data from google drive

```
[R] from google.colab import drive
drive.mount('/content/drive')

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).
```

70°F Party cloudy

Google
Untitled9.ipynb - Colaboratory
Data science Project - Google Drive

https://colab.research.google.com/drive/1QweoX79kHwSp56_h0QOG2SLvNMN18/#scrollTo=ZEq5C9_dNYE

Import favorites | ASUS Software Port... | MyASUS Software... | McAfee LiveSafe | Can't sign up my d... | RPCS3 - Compatibil... | New tabs

File Edit View Insert Runtime Tools Help

+ Code + Text

```
[ ] + Image Augmentation

[ ]
datagen = ImageDataGenerator(rotation_range = 360,
                             width_shift_range = 0.2,
                             height_shift_range = 0.2,
                             zoom_range = 0.2,
                             horizontal_flip = True,
                             vertical_flip = True)
```

SPLITTING DATA

70°F Party cloudy

Google
Untitled9.ipynb - Col... | Colab Notebooks | Untitled9.ipynb - Col... | Welcome To Colab... | Untitled10.ipynb | Data science Project

https://drive.google.com/drive/folders/1b0bX3p9pBY7M5mccskfAtsyp9Lxw2

Import favorites | ASUS Software Port... | MyASUS Software... | McAfee LiveSafe | Can't sign up my d... | RPCS3 - Compatibil... | New tabs

Drive

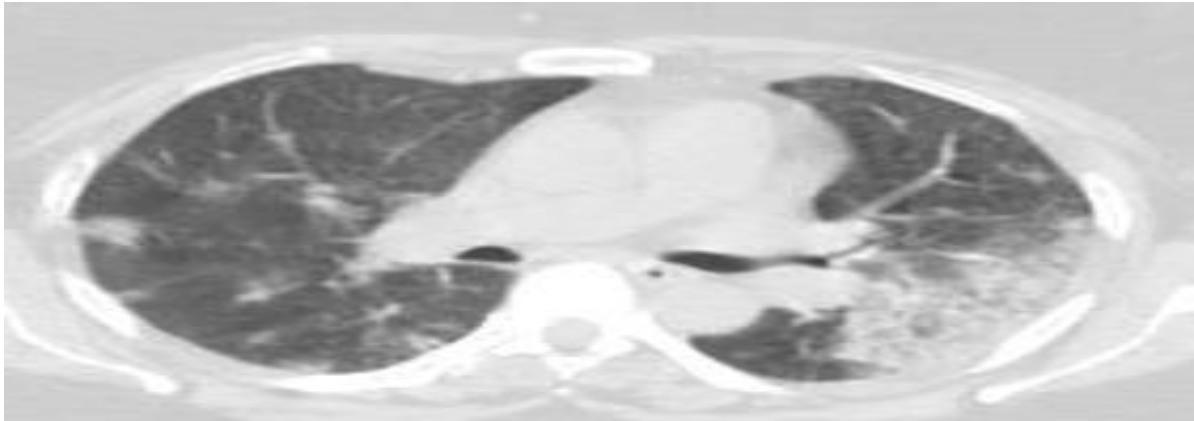
Search in Drive

My Drive > Colab Notebooks > Data science Project

| Name | Owner | Last modified | File size |
|-----------|-------|---------------|-----------|
| COVID | me | 2:16 AM | — |
| non-COVID | me | 2:16 AM | — |

Storage
5.45 GB of 15 GB used
Buy storage

Uploading 2 items
18 min left...
COVID 0 of 1252
non-COVID 393 of 1234



1 – The images are in different sizes so you have to take a fixed size on which you have to work-

We can use the Python Imaging Library (PIL) to resize images where we have to work.

First, import the PIL library:

```
import PIL
```

Then, open the image you want to resize.

```
Img = PIL.Image.open("image_name.jpg")
```

Now you can set the desired size for the image:

```
new_width = 200 new_height = 300
```

Finally, you can use the resize function of the PIL library to resize the image:

```
img.resize((new_width, new_height))
```

The image is now resized to the desired size.

SPLITTING DATA

The data can be split into train and test sets using the `train_test_split()` method from the `scikit-learn` library. This method takes two parameters: `X`, which is the feature set, and `y`, which is the target variable. The `test_size` parameter is used to specify the proportion of the data that should be split into the test set. The `random_state` parameter is used to ensure that the same split is used each time the code is run.

Bulding and Visualizing the model using ResNet-50

ResNet-Building50 is a convolutional neural network that is trained on more than a million images from the ImageNet database. It is a 50-layer deep convolutional network and is a variant of the ResNet architecture. ResNet-50 can be used for image classification, object detection, and other computer vision tasks.

In this example, ResNet-50 will be used to build and visualize a model for image classification. The code below imports the necessary libraries, loads the ResNet-50 model, and prepares the data for training.

END