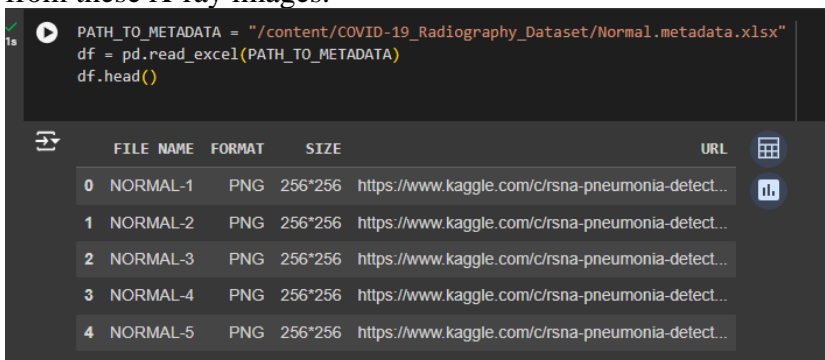
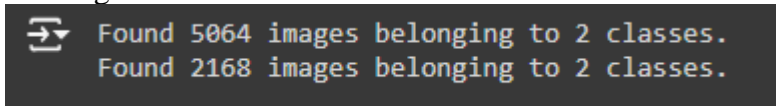


## Data Collection and Preprocessing Phase

Date	8th July 2024
Team ID	SWTID1720195938
Project Title	<b>CovidVision: Advanced COVID-19 Detection from Lung X-Rays with Deep Learning</b>
Maximum Marks	6 Marks

Section	Description																														
Data Overview	<p>Give an overview of the data, which you are going to use in your project. The dataset consists of lung X-ray images, including COVID-19-positive cases, normal lung images, lung opacity images, and viral pneumonia images. The images are in PNG format with a resolution of 299x299 pixels. The objective is to use deep learning models to detect COVID-19 infections from these X-ray images.</p>  <pre>PATH_TO_METADATA = "/content/COVID-19_Radiography_Dataset/Normal.metadata.xlsx" df = pd.read_excel(PATH_TO_METADATA) df.head()</pre> <table><thead><tr><th></th><th>FILE NAME</th><th>FORMAT</th><th>SIZE</th><th>URL</th></tr></thead><tbody><tr><td>0</td><td>NORMAL-1</td><td>PNG</td><td>256*256</td><td>https://www.kaggle.com/c/rsna-pneumonia-detect...</td></tr><tr><td>1</td><td>NORMAL-2</td><td>PNG</td><td>256*256</td><td>https://www.kaggle.com/c/rsna-pneumonia-detect...</td></tr><tr><td>2</td><td>NORMAL-3</td><td>PNG</td><td>256*256</td><td>https://www.kaggle.com/c/rsna-pneumonia-detect...</td></tr><tr><td>3</td><td>NORMAL-4</td><td>PNG</td><td>256*256</td><td>https://www.kaggle.com/c/rsna-pneumonia-detect...</td></tr><tr><td>4</td><td>NORMAL-5</td><td>PNG</td><td>256*256</td><td>https://www.kaggle.com/c/rsna-pneumonia-detect...</td></tr></tbody></table>		FILE NAME	FORMAT	SIZE	URL	0	NORMAL-1	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...	1	NORMAL-2	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...	2	NORMAL-3	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...	3	NORMAL-4	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...	4	NORMAL-5	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...
	FILE NAME	FORMAT	SIZE	URL																											
0	NORMAL-1	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...																											
1	NORMAL-2	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...																											
2	NORMAL-3	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...																											
3	NORMAL-4	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...																											
4	NORMAL-5	PNG	256*256	https://www.kaggle.com/c/rsna-pneumonia-detect...																											
Resizing	<p>Images will be resized to 299x299 pixels to ensure uniformity and compatibility with the input requirements of the deep learning models.</p>  <pre>Found 5064 images belonging to 2 classes. Found 2168 images belonging to 2 classes.</pre>																														
Normalization & Data Augmentation	<p>Pixel values of the images will be normalized to a range of [0, 1] to facilitate faster convergence during training. Data augmentation techniques, such as random rotation, flipping, shifting, zooming, and shearing, will be applied to increase the</p>																														

	diversity of the training set and improve the model's generalization.
Batch Normalization	Batch normalization will be applied to the input of each layer in the neural network to stabilize and speed up the training process.
<b>Data Preprocessing Code Screenshots</b>	
Loading Data	<pre> 74 files.upload() [4] kaggle.json + kaggle.json(application/json) - 66 bytes, last modified: 7/9/2024 - 100% done Saving kaggle.json to kaggle.json {'kaggle.json': b'{"username": "Tanishaj04", "key": "a3cb43368a9908ff3cdeca529146258d"}'}</pre> <pre> [5] mkdir -p ~/.kaggle [cp kaggle.json ~/.kaggle]</pre> <pre> [6] !kaggle datasets download -d tawefurrahman/covid19-radiography-database Warning: Your Kaggle API key is readable by other users on this system! To fix this, you can run 'chmod 600 /root/.kaggle/kaggle.json' Dataset URL: https://www.kaggle.com/datasets/tawefurrahman/covid19-radiography-database License(s): copyright, authors Downloading covid19-radiography-database.zip to /content 100% 772M/772M [00:05&lt;00:00, 160MB/s] 100% 772M/772M [00:05&lt;00:00, 160MB/s]</pre> <pre> 76 !unzip '/content/covid19-radiography-database.zip' Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-811.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-812.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-813.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-814.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-815.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-816.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-817.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-818.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-819.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-820.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-821.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-822.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-823.png Inflating: COVID-19_Radiography_Dataset/Viral_Pneumonia/masks/Viral_Pneumonia-824.png</pre>
Resizing	<pre> [0] IMAGE_SIZE = (299,299) IMAGE_SHAPE = IMAGE_SIZE + (3,)  [20] from tensorflow.keras.preprocessing.image import ImageDataGenerator img_height, img_width= IMAGE_SIZE batch_size=32 train_datagen = ImageDataGenerator(validation_split=0.3)  train_generator = train_datagen.flow_from_directory(     train_data_dir,     target_size=(img_height, img_width),     batch_size=batch_size,     class_mode='binary',     subset='training')  validation_generator = train_datagen.flow_from_directory(     train_data_dir,     target_size=(img_height, img_width),     batch_size=batch_size,     class_mode='binary',     subset='validation')</pre> <pre> Found 5064 images belonging to 2 classes. Found 2168 images belonging to 2 classes.</pre>

<p>Normalization &amp; Data Augmentation</p>	<pre>Found 5064 images belonging to 2 classes. Found 2168 images belonging to 2 classes.  train_datagen = ImageDataGenerator(     rescale=1./255,     rotation_range=20,     width_shift_range=0.2,     height_shift_range=0.2,     shear_range=0.2,     zoom_range=0.2,     horizontal_flip=True,     fill_mode='nearest',     validation_split=0.3)</pre>
<p>Batch Normalization</p>	<pre>from keras.models import load_model from keras.layers import BatchNormalization from keras.utils import register_keras_serializable # Import from the correct module  # Define the custom BatchNormalization layer class @register_keras_serializable() # Register the custom layer class CustomBatchNormalization(BatchNormalization):     def __init__(self, **kwargs):         super(CustomBatchNormalization, self).__init__(**kwargs)      @classmethod     def from_config(cls, config):         config['axis'] = config.get('axis', []) if isinstance(config.get('axis', []), list) else config.get('axis')         return cls(**config)  # Load the model with the custom objects model = load_model('Xception-Covid.h5', custom_objects={'CustomBatchNormalization': CustomBatchNormalization})  # Optionally, save the model in the new Keras format model.save('Xception-Covid.keras')  # Load the model in the new Keras format model = load_model('Xception-Covid.keras')</pre>