

## Data Collection and Preprocessing Phase

Date	01 May 2025
Team ID	739942
Project Title	CovidVision: Advanced COVID-19 Detection From Lung X-Rays With Deep Learning Using IBM Cloud
Maximum Marks	6 Marks

### Preprocessing Template

The images will be preprocessed by resizing, normalizing, augmenting, denoising, adjusting contrast, detecting edges, converting color space, cropping, batch normalizing, and whitening data. These steps will enhance data quality, promote model generalization, and improve convergence during neural network training, ensuring robust and efficient performance across various computer vision tasks.

Section	Description
Data Overview	Overview of the dataset is sourced from <b>kaggle</b> , consisting of chest X-ray images to train and evaluate its deep learning model for detecting COVID-19 cases.
Resizing	Images are resized to a target size of <b>224x224 pixels</b>
Normalization	Normalize pixel values to a range of <b>[0,1]</b> by dividing by 255.
Data Augmentation	Augmentation techniques such as flipping, zooming, and shearing are applied enhance the dataset.

## Data Preprocessing Code Screenshots

### Loading Data

```
import kagglehub

# Download latest version
path = kagglehub.dataset_download("tausifurrahman/covid19-radiography-database")

print("Path to dataset files:", path)

Downloading from https://www.kaggle.com/api/v1/datasets/download/tausifurrahman/covid19-radiography-database?dataset_version_number=5...
100%|#####| 778M/778M [00:05<00:00, 156MB/s]extracting files...

Path to dataset files: /root/.cache/kagglehub/datasets/tausifurrahman/covid19-radiography-database/versions/5
```

### Resizing

```
# Use the 'path' variable obtained from kagglehub.dataset_download()
training_generator = data_gen.flow_from_directory(
    directory=path, # Use the correct path
    target_size=(299, 299),
    batch_size=32,
    class_mode='categorical'
)

# Use the 'path' variable for the test generator as well
test_generator = data_gen.flow_from_directory(
    directory=path, # Use the correct path
    target_size=(299, 299),
    batch_size=32,
    class_mode='categorical'
)
```

### Normalization

```
# Define the ImageDataGenerator
data_gen = ImageDataGenerator(
    rescale=1./255,
    validation_split=0.2, # Use 20% of the data for validation
```

### Data Augmentation

```
from tensorflow.keras.preprocessing.image import ImageDataGenerator

# Define the ImageDataGenerator
data_gen = ImageDataGenerator(
    rescale=1./255,
    validation_split=0.2, # Use 20% of the data for validation
    rotation_range=40,
    width_shift_range=0.2,
    height_shift_range=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True,
    fill_mode='nearest'
)
```

