Data Collection and Preprocessing Phase

Date	12 July 2024
Team ID	SWTID1720157891
Project Title	Rice Classification using CNN
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	For this project, we are utilizing a dataset consisting of various rice grain images, categorized into different types. The dataset is structured in directories where each sub-directory represents a class of rice. This data is crucial for training our convolutional neural network (CNN) to distinguish between different types of rice grains accurately.
Resizing	Resizing the images to a uniform size is essential for consistent input to the CNN. In our case, we resized all images to 128x128 pixels.
Normalization	Normalization is applied to scale pixel values to the range [0, 1], which helps in faster convergence of the neural network during training.
Data Augmentation	Data augmentation techniques are employed to increase the diversity of the training set and improve the model's robustness. Techniques used include: Horizontal flipping Shearing Zooming

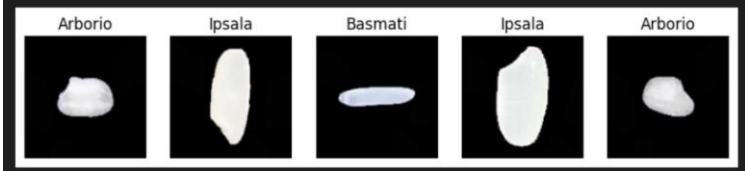
Denoising	Applying denoising filters can help in reducing noise in the images, which can improve model performance.
Edge Detection	Edge detection algorithms highlight prominent edges in the images, which can be useful for emphasizing structural features
Color Space Conversion	Conversion between color spaces can be useful for different image processing tasks. For example, converting RGB images to grayscale can reduce the dimensionality of the input data
Image Cropping	Cropping images to focus on regions containing objects of interest can improve the model's ability to learn relevant features.
Batch Normalization	Batch normalization helps to stabilize and accelerate the training of deep neural networks.
Data Preprocessing Code Screenshots	

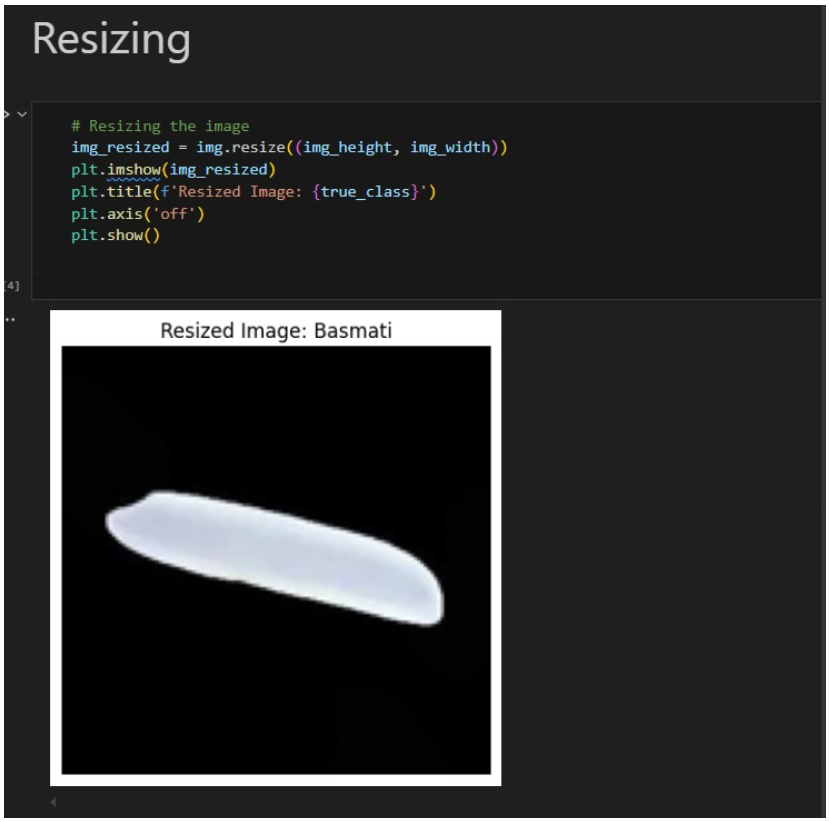
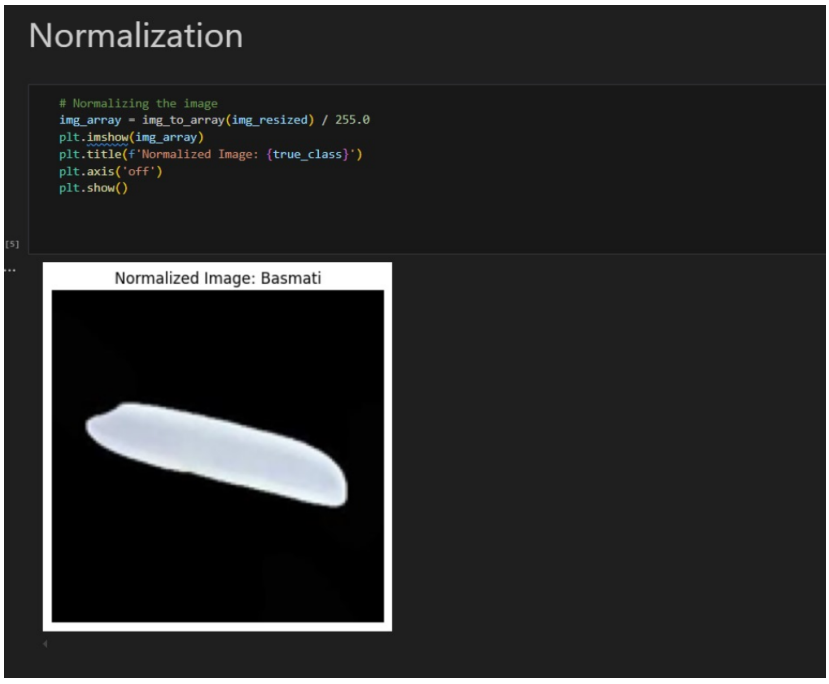
Loading Data	Loading the data
	`import os`
	`import numpy as np`
	`import random`
	`import matplotlib.pyplot as plt`
	`import tensorflow as tf`
	`from tensorflow.keras.preprocessing.image import ImageDataGenerator, load_img, img_to_array`
	`from tensorflow.keras.models import Sequential`
	`from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense, Dropout, BatchNormalization`
	`from sklearn.metrics import classification_report, confusion_matrix`
	`from tensorflow.keras.preprocessing import image`
	`# Path to the dataset directory`
	`dataset_dir = r'C:\Users\Gowtham Ravuri\OneDrive\Desktop\Rice Classification Project\Rice_Image_Dataset'  # Using raw string`
	`# Parameters`
	`img_height, img_width = 128, 128`
	`batch_size = 32`
	`epochs = 5`

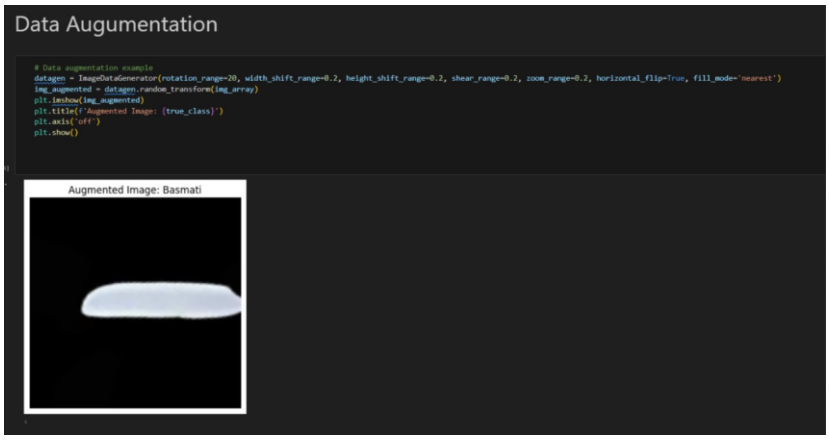
```
# Data generators for training and validation with data augmentation
datagen = ImageDataGenerator(
    rescale=1./255,
    validation_split=0.2,
    shear_range=0.2,
    zoom_range=0.2,
    horizontal_flip=True
)

train_generator = datagen.flow_from_directory(
    dataset_dir,
    target_size=(img_height, img_width),
    batch_size=batch_size,
    class_mode='categorical',
    subset='training'
)

validation_generator = datagen.flow_from_directory(
    dataset_dir,
    target_size=(img_height, img_width),
    batch_size=batch_size,
    class_mode='categorical',
    subset='validation'
)

# Preview a few images with labels
def preview_images(generator, num_images=5):
    class_names = list(generator.class_indices.keys())
    images, labels = next(generator)
    plt.figure(figsize=(10, 10))
    for i in range(num_images):
        ax = plt.subplot(1, num_images, i + 1)
        plt.imshow(images[i])
        plt.title(class_names[np.argmax(labels[i])])
        plt.axis("off")
    plt.show()

# Preview images from the training set
preview_images(train_generator)
```

Output-

```
Found 60000 images belonging to 5 classes.
Found 15000 images belonging to 5 classes.
```



Resizing	
Normalization	

Data Augmentation	
Denoising	

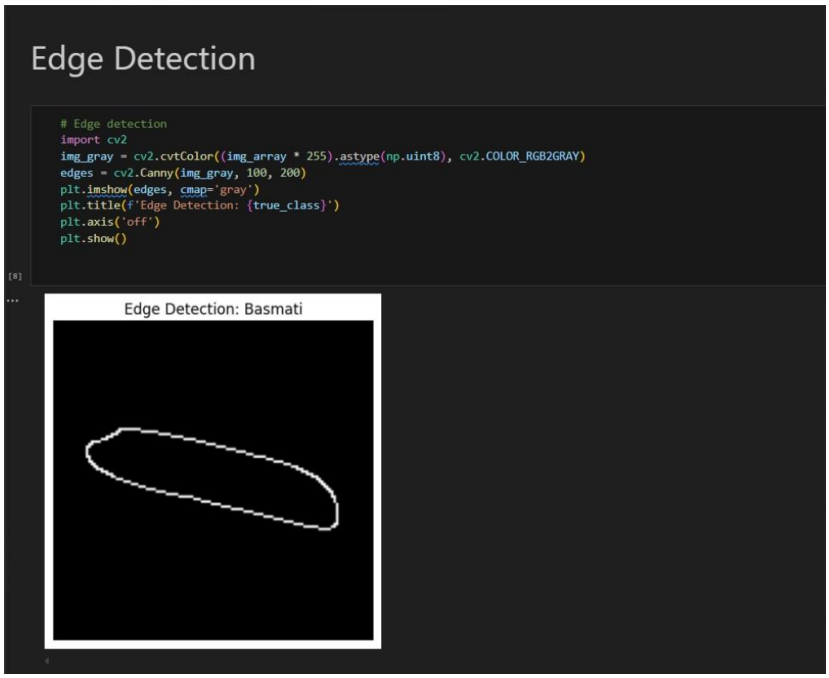
Edge Detection	
Color Space Conversion	

Image Cropping	
Batch Normalization