Data Collection and Preprocessing Phase

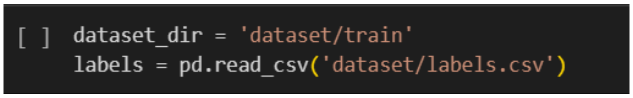
Date	11 July 2024
Team ID	SWTID1720067113
Project Title	Dog Breed Identification using Transfer Learning
Maximum Marks	6 Marks

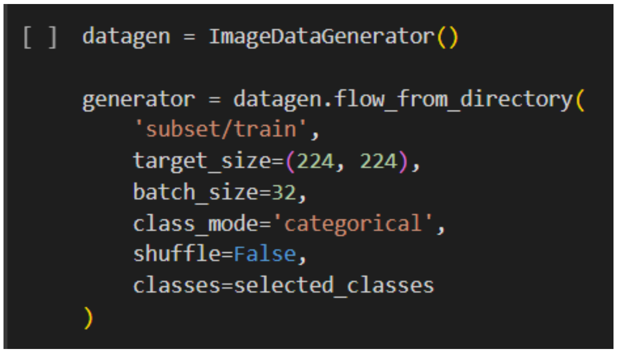
Preprocessing Template

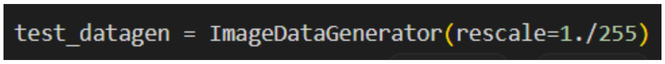
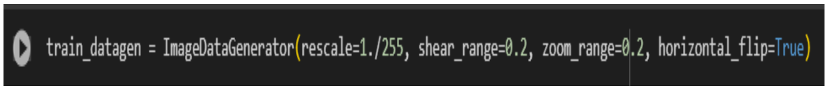
The images will be preprocessed by resizing, normalizing, augmenting, denoising, adjusting contrast, detecting edges, converting color space, cropping, batch normalizing, and whitening data. These steps will enhance data quality, promote model generalization, and improve convergence during neural network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

Section	**Description**
Data Overview	The dataset used in this project comprises images of dogs labelled with their corresponding breeds. The primary dataset is sourced from the Kaggle Dog Breed Identification competition. It includes a diverse collection of dog images spanning multiple breeds, which will be used for training and evaluating the dog breed identification model. **Dimension:** • **Training Set:** 10,222 images • **Test Set:** 10,357 images • **Number of Classes:** 120 dog breeds • **Feature Dimension:** Images resized to 224x224 pixels with 3 color channels (RGB)
Resizing	Resize images to a specified target size.
Normalization	Normalize pixel values to a specific range.
Data Augmentation	Apply augmentation techniques such as flipping, rotation, shifting, zooming, or shearing.

Denoising	Apply denoising filters to reduce noise in the images.
Edge Detection	Apply edge detection algorithms to highlight prominent edges in the images.
Color Space Conversion	Convert images from one color space to another.
Image Cropping	Crop images to focus on the regions containing objects of interest.
Batch Normalization	Apply batch normalization to the input of each layer in the neural network.
Data Preprocessing Code Screenshots	

Loading Data	
	`[ ] dataset_dir = 'dataset/train'`
	`labels = pd.read_csv('dataset/labels.csv')`

Resizing	
	`[ ] datagen = ImageDataGenerator()`
	
	`generator = datagen.flow_from_directory(`
	`    'subset/train',`
	`    target_size=(224, 224),`
	`    batch_size=32,`
	`    class_mode='categorical',`
	`    shuffle=False,`
	`    classes=selected_classes`
	`)`

Normalization	 `test_datagen = ImageDataGenerator(rescale=1./255)`
Data Augmentation	 `train_datagen = ImageDataGenerator(rescale=1./255, shear_range=0.2, zoom_range=0.2, horizontal_flip=True)`