

## Data Collection and Preprocessing Phase

Date	20 June 2025
Team ID	SWTID1749826875
Project Title	Dog Breed Identification using Transfer Learning
Maximum Marks	6 Marks

### Preprocessing

The images will be preprocessed by resizing, normalizing, converting color space and batch normalizing. 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 contains 10,222 images categorized into 120 classes. For training purposes, only 20 of these classes were selected, comprising a total of 1,683 images.
Resizing	The images were resized to $224 \times 224$ pixels to ensure uniformity.
Normalization	Pixel values were normalized to a target range of 0 to 1.
Color Space Conversion	The images are taken BGR format
Batch Normalization	A batch size of 32 was used during model training.

<h2>Data Preprocessing Code Screenshots</h2>	
<h2>Loading Data</h2>	<pre>[ ] lrm -rf /content/subset  [ ] pip install -q kaggle  [ ] mkdir ~/.kaggle  [ ] cp kaggle.json ~/.kaggle  [ ] kaggle competitions download -c dog-breed-identification  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' Downloading dog-breed-identification.zip to /content 96% 645M/691M [00:01&lt;00:00, 263MB/s] 100% 691M/691M [00:01&lt;00:00, 379MB/s]  [ ] unzip /content/dog-breed-identification.zip  inflating: train/87a2b41bedf49388251c98897885887.jpg inflating: train/87a8f6b9a8e618486e7662a133e9e61.jpg inflating: train/87b33815ce34f2c2d98b2374c32ce3a5.jpg inflating: train/87c3889fb8c7c00ae4c78b939538f84.jpg</pre>
<h2>Resizing</h2>	<pre># datagen = ImageDataGenerator () generator = train_datagen.flow_from_directory(     '/content/subset/train',     target_size=(224, 224), # Adjust target size as needed     batch_size=32,     class_mode='categorical',     shuffle=False, # Ensure order is maintained for class indices     classes=selected_classes # Specify the selected classes )</pre>
<h2>Normalization</h2>	<pre>[ ] #import image datagenerator library from tensorflow.keras.preprocessing.image import ImageDataGenerator train_datagen=ImageDataGenerator(rescale=1./255, shear_range=0.2, zoom_range=0.2, horizontal_flip=True)</pre>
<h2>Batch Normalization</h2>	<pre># datagen = ImageDataGenerator () generator = train_datagen.flow_from_directory(     '/content/subset/train',     target_size=(224, 224), # Adjust target size as needed     batch_size=32,     class_mode='categorical',     shuffle=False, # Ensure order is maintained for class indices     classes=selected_classes # Specify the selected classes )</pre>