```
!unzip "/content/FruitinAmazon.zip"
```

```
inflating: FruitinAmazon/test/guarana/download (5).jpeg
inflating: FruitinAmazon/train/graviola/download (6).jpeg
inflating: FruitinAmazon/train/graviola/images (3).jpeg
inflating: FruitinAmazon/train/graviola/images (10).jpeg
inflating: FruitinAmazon/train/pupunha/images (12).jpeg
inflating: FruitinAmazon/train/graviola/images (8).jpeg
inflating: FruitinAmazon/train/guarana/download (10).jpeg
inflating: FruitinAmazon/train/tucuma/images (2).jpeg
inflating: FruitinAmazon/train/tucuma/images (1).jpeg
inflating: FruitinAmazon/train/guarana/images (6).jpeg
inflating: FruitinAmazon/train/guarana/images (5).jpeg
inflating: FruitinAmazon/train/guarana/download (1).jpeg
inflating: FruitinAmazon/train/guarana/images (2).jpeg
inflating: FruitinAmazon/train/tucuma/download (9).jpeg
inflating: FruitinAmazon/train/graviola/download (8).jpeg
inflating: FruitinAmazon/train/pupunha/images (11).jpeg
inflating: FruitinAmazon/train/acai/images (6).jpeg
inflating: FruitinAmazon/test/guarana/images (4).jpeg
inflating: FruitinAmazon/train/graviola/download (5).jpeg
inflating: FruitinAmazon/train/tucuma/images.jpeg
inflating: FruitinAmazon/train/tucuma/images (9).jpeg
inflating: FruitinAmazon/train/graviola/download.jpeg
inflating: FruitinAmazon/train/graviola/images (5).jpeg
inflating: FruitinAmazon/train/graviola/images (2).jpeg
inflating: FruitinAmazon/train/guarana/download (9).jpeg
inflating: FruitinAmazon/train/acai/images (12).jpeg
inflating: FruitinAmazon/train/tucuma/images (3).jpeg
inflating: FruitinAmazon/train/tucuma/images (7).jpeg
inflating: FruitinAmazon/train/graviola/images.jpeg
inflating: FruitinAmazon/train/tucuma/images (5).jpeg
inflating: FruitinAmazon/train/tucuma/download (3).jpeg
inflating: FruitinAmazon/train/cupuacu/images (4).jpeg
inflating: FruitinAmazon/train/cupuacu/images (7).jpeg
inflating: FruitinAmazon/train/cupuacu/download.jpeg
inflating: FruitinAmazon/test/acai/images (17).jpeg
inflating: FruitinAmazon/train/graviola/images (1).jpeg
inflating: FruitinAmazon/train/cupuacu/images (13).jpeg
inflating: FruitinAmazon/train/cupuacu/images (12).jpeg
inflating: FruitinAmazon/train/tucuma/download (6).jpeg
inflating: FruitinAmazon/train/cupuacu/images (5).jpeg
inflating: FruitinAmazon/train/guarana/images (7).jpeg
inflating: FruitinAmazon/train/tucuma/images (6).jpeg
inflating: FruitinAmazon/train/cupuacu/images (6).jpeg
inflating: FruitinAmazon/train/tucuma/images (8).jpeg
inflating: FruitinAmazon/train/cupuacu/download (1).jpeg
inflating: FruitinAmazon/test/graviola/download (4).jpeg
inflating: FruitinAmazon/train/tucuma/images (4).jpeg
inflating: FruitinAmazon/train/cupuacu/images (8).jpeg
inflating: FruitinAmazon/train/cupuacu/images (1).jpeg
inflating: FruitinAmazon/train/cupuacu/images (10).jpeg
inflating: FruitinAmazon/train/tucuma/download (7).jpeg
inflating: FruitinAmazon/train/guarana/download.jpeg
inflating: FruitinAmazon/train/cupuacu/images (2).jpeg
inflating: FruitinAmazon/train/graviola/images (9).jpeg
inflating: FruitinAmazon/train/cupuacu/images (11).jpeg
inflating: FruitinAmazon/train/cupuacu/images (9).jpeg
inflating: FruitinAmazon/train/tucuma/download.jpeg
inflating: FruitinAmazon/train/cupuacu/images.jpeg
```

6 Simple CNN Implemented using Keras.

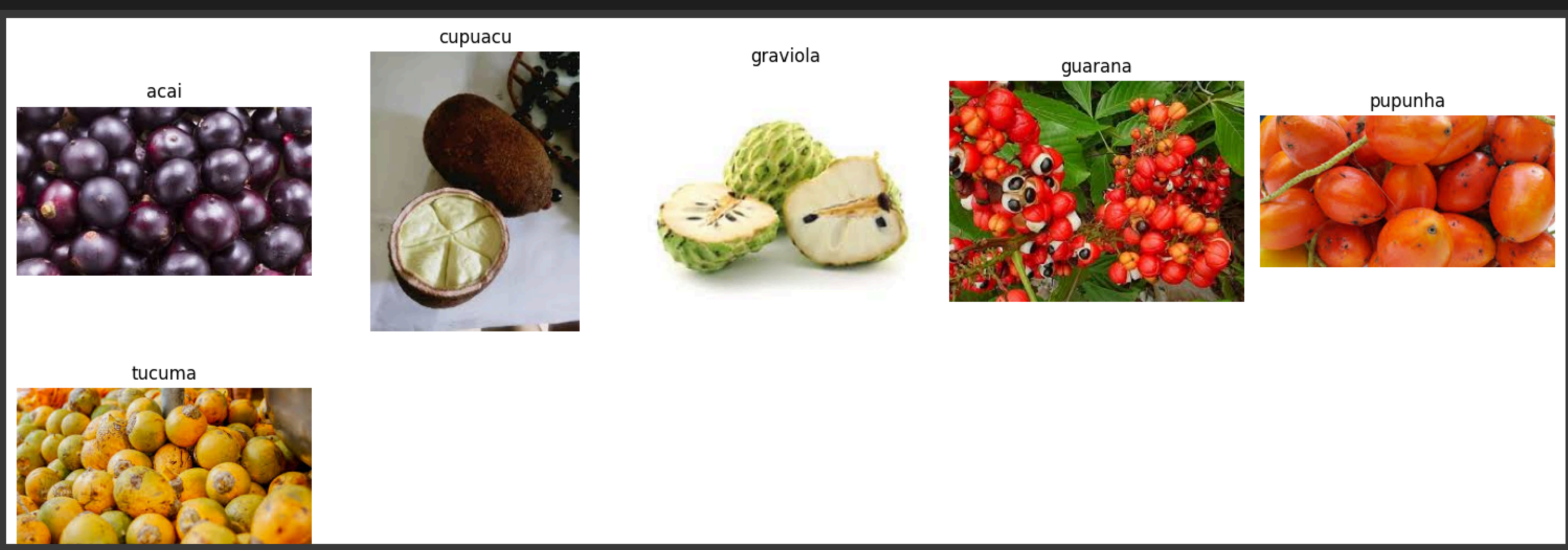
```
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers
import numpy as np
# Load a sample dataset (MNIST for simplicity)
(x_train, y_train), (x_test, y_test) = keras.datasets.mnist.load_data()
# Normalize and reshape data
x_train = x_train.astype("float32") / 255.0
x_test = x_test.astype("float32") / 255.0
x_train = np.expand_dims(x_train, axis=-1) # Add channel dimension
x_test = np.expand_dims(x_test, axis=-1)
# Define a simple CNN model
model = keras.Sequential([
layers.Conv2D(32, (3, 3), activation="relu", input_shape=(28, 28, 1)),
layers.MaxPooling2D((2, 2)),
layers.Conv2D(64, (3, 3), activation="relu"),
layers.MaxPooling2D((2, 2)),
layers.Flatten(),
layers.Dense(128, activation="relu"),
layers.Dense(10, activation="softmax") # 10 classes for MNIST digits
])
# Compile the model
model.compile(optimizer="adam",
loss="sparse_categorical_crossentropy",
metrics=["accuracy"])
# Train the model
model.fit(x_train, y_train, epochs=5, batch_size=32, validation_data=(x_test, y_test))
# Evaluate the model
test_loss, test_acc = model.evaluate(x_test, y_test)
print(f"Test accuracy: {test_acc:.4f}")
# Make predictions
predictions = model.predict(x_test[:5])
predicted_labels = np.argmax(predictions, axis=1)
print("Predicted labels:", predicted_labels)
print("Actual labels: ", y_test[:5])
```

```
Downloading data from https://storage.googleapis.com/tensorflow/tf-keras-datasets/mnist.npz
11490434/11490434 ━━━━━━━━━━━━━━━━━━━━ 0s 0us/step
/usr/local/lib/python3.11/dist-packages/keras/src/layers/convolutional/base_conv.py:107: UserWarning: Do not pass an `input_shape`/`input_dim` argument to a layer. When using Sequential models, prefer using an `Input(shape)` object as the first layer in the model instead.
  super().__init__(activity_regularizer=activity_regularizer, **kwargs)
Epoch 1/5
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 12s 4ms/step - accuracy: 0.9112 - loss: 0.2966 - val_accuracy: 0.9846 - val_loss: 0.0471
Epoch 2/5
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 6s 3ms/step - accuracy: 0.9862 - loss: 0.0445 - val_accuracy: 0.9895 - val_loss: 0.0290
Epoch 3/5
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 11s 4ms/step - accuracy: 0.9912 - loss: 0.0284 - val_accuracy: 0.9896 - val_loss: 0.0319
Epoch 4/5
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 6s 3ms/step - accuracy: 0.9936 - loss: 0.0192 - val_accuracy: 0.9910 - val_loss: 0.0256
Epoch 5/5
1875/1875 ━━━━━━━━━━━━━━━━━━━━ 10s 3ms/step - accuracy: 0.9959 - loss: 0.0129 - val_accuracy: 0.9907 - val_loss: 0.0285
313/313 ━━━━━━━━━━━━━━━━━━━━ 1s 2ms/step - accuracy: 0.9876 - loss: 0.0364
Test accuracy: 0.9907
1/1 ━━━━━━━━━━━━━━━━━━━━ 1s 618ms/step
Predicted labels: [7 2 1 0 4]
Actual labels:  [7 2 1 0 4]
```

7 Exercise.

Task 1: Data Understanding and Visualization:

```
import os
import zipfile
import random
import matplotlib.pyplot as plt
import matplotlib.image as mpimg

# Path to the ZIP file
zip_path = "/content/FruitinAmazon.zip"  # Update this if needed
extract_path = "/content/FruitinAmazon"

# Extract the ZIP file if not already extracted
if not os.path.exists(extract_path):
    with zipfile.ZipFile(zip_path, 'r') as zip_ref:
        zip_ref.extractall(extract_path)

# Define the train dataset directory
train_dir = os.path.join(extract_path, "train")  # Adjust if needed

# Check if the directory exists
if not os.path.isdir(train_dir):
    raise ValueError(f"Train directory not found: {train_dir}")

# Get the list of class directories
class_names = sorted([d for d in os.listdir(train_dir) if os.path.isdir(os.path.join(train_dir, d))])

# Select one random image from each class
selected_images = []
labels = []

for class_name in class_names:
    class_path = os.path.join(train_dir, class_name)
    images = [f for f in os.listdir(class_path) if os.path.isfile(os.path.join(class_path, f))]

    if images:
        img_path = os.path.join(class_path, random.choice(images))
        selected_images.append(img_path)
        labels.append(class_name)

# Plot images in a grid format
num_classes = len(selected_images)
cols = 5  # Number of columns
rows = (num_classes + cols - 1) // cols  # Compute rows dynamically

fig, axes = plt.subplots(rows, cols, figsize=(15, 6))
axes = axes.flatten()

for i in range(rows * cols):
    if i < num_classes:
        img = mpimg.imread(selected_images[i])
        axes[i].imshow(img)
        axes[i].set_title(labels[i])
        axes[i].axis("off")
    else:
        axes[i].axis("off")  # Hide extra subplots

plt.tight_layout()
plt.show()
```



• What did you Observe?

-> Class Distribution, Image Variety, Dataset Size and Balance, Image Quality & Resolution, Potential Preprocessing Needs.

Check for Corrupted Image

```
import os
from PIL import Image

# Define the path to the train dataset folder
train_dir = "/content/FruitinAmazon/train"  # Update this path if necessary

# List to store corrupted image paths
corrupted_images = []

# Iterate through each class directory
for class_name in sorted(os.listdir(train_dir)):
```

```
    class_path = os.path.join(train_dir, class_name)

    # Ensure it's a directory
    if os.path.isdir(class_path):
        for image_name in os.listdir(class_path):
            image_path = os.path.join(class_path, image_name)

            try:
                # Try opening the image
                with Image.open(image_path) as img:
                    img.verify()  # Verify the image integrity

            except (IOError, SyntaxError):
                # If an error occurs, remove the corrupted image
                corrupted_images.append(image_path)
                os.remove(image_path)
                print(f"Removed corrupted image: {image_path}")

# Final Report
if not corrupted_images:
    print("No corrupted images found.")
else:
    print(f"\nTotal corrupted images removed: {len(corrupted_images)}")
```

```
No corrupted images found.
```

Task 2: Loading and Preprocessing Image Data in keras:

```
import tensorflow as tf

# Define dataset directory
train_dir = "/content/FruitinAmazon/train"  # Update path if necessary

# Define image size and batch size
img_height = 128  # Target image height
img_width = 128   # Target image width
batch_size = 32   # Number of images per batch
validation_split = 0.2  # 80% training, 20% validation

# Create a preprocessing layer for normalization
rescale = tf.keras.layers.Rescaling(1./255)  # Normalize pixel values to [0, 1]

# Load training dataset
train_ds = tf.keras.preprocessing.image_dataset_from_directory(
    train_dir,
    labels='inferred',  # Uses subdirectory names as class labels
    label_mode='int',   # Labels are encoded as integers
    image_size=(img_height, img_width),  # Resize images
    interpolation='nearest',
    batch_size=batch_size,
    shuffle=True,  # Shuffle training data
    validation_split=validation_split,  # Split data
    subset='training',  # Use training subset
    seed=123  # Ensure reproducibility
)

# Apply normalization to training dataset
train_ds = train_ds.map(lambda x, y: (rescale(x), y))

# Load validation dataset
val_ds = tf.keras.preprocessing.image_dataset_from_directory(
    train_dir,
    labels='inferred',
    label_mode='int',
    image_size=(img_height, img_width),
    interpolation='nearest',
    batch_size=batch_size,
    shuffle=False,  # No need to shuffle validation data
    validation_split=validation_split,
    subset='validation',  # Use validation subset
    seed=123
)

# Apply normalization to validation dataset
val_ds = val_ds.map(lambda x, y: (rescale(x), y))

# Print dataset info
print(f"Training batches: {len(train_ds)}, Validation batches: {len(val_ds)}")
```

```
Found 90 files belonging to 6 classes.
Using 72 files for training.
Found 90 files belonging to 6 classes.
Using 18 files for validation.
Training batches: 3, Validation batches: 1
```

Task 3 - Implement a CNN

```
import tensorflow as tf
from tensorflow import keras
from tensorflow.keras import layers

# Define image size and number of classes
img_height = 128
img_width = 128
num_classes = 10  # Update this based on the dataset

# Build the CNN model
model = keras.Sequential([
    # Convolutional Layer 1
    layers.Conv2D(filters=32, kernel_size=(3, 3), padding="same", strides=1, activation="relu",
                  input_shape=(img_height, img_width, 3)),
    layers.MaxPooling2D(pool_size=(2, 2), strides=2),

    # Convolutional Layer 2
    layers.Conv2D(filters=32, kernel_size=(3, 3), padding="same", strides=1, activation="relu"),
    layers.MaxPooling2D(pool_size=(2, 2), strides=2),

    # Flatten Layer
    layers.Flatten(),

    # Fully Connected Layers
    layers.Dense(64, activation="relu"),
    layers.Dense(128, activation="relu"),

    # Output Layer
    layers.Dense(num_classes, activation="softmax")  # Softmax for multi-class classification
])

# Compile the model
model.compile(optimizer="adam",
              loss="sparse_categorical_crossentropy",
              metrics=["accuracy"])

# Print the model summary
model.summary()
```

Model: "sequential_1"

Layer (type)	Output Shape	Param #
conv2d_2 (Conv2D)	(None, 128, 128, 32)	896
max_pooling2d_2 (MaxPooling2D)	(None, 64, 64, 32)	0
conv2d_3 (Conv2D)	(None, 64, 64, 32)	9,248
max_pooling2d_3 (MaxPooling2D)	(None, 32, 32, 32)	0
flatten_1 (Flatten)	(None, 32768)	0
dense_2 (Dense)	(None, 64)	2,097,216
dense_3 (Dense)	(None, 128)	8,320
dense_4 (Dense)	(None, 10)	1,290

```
Total params: 2,116,970 (8.08 MB)
Trainable params: 2,116,970 (8.08 MB)
Non-trainable params: 0 (0.00 B)
```

Task 4: Compile the Model

```
from tensorflow.keras.callbacks import ModelCheckpoint, EarlyStopping

# Compile the model
model.compile(
    optimizer="adam",
    loss="sparse_categorical_crossentropy",
    metrics=["accuracy"]
)

# Define callbacks
checkpoint_cb = ModelCheckpoint(
    "best_model.h5", monitor="val_accuracy", save_best_only=True, verbose=1
)
early_stopping_cb = EarlyStopping(
    monitor="val_loss", patience=10, restore_best_weights=True, verbose=1
)

# Train the model
history = model.fit(
    train_ds,
    validation_data=val_ds,
    batch_size=16,  # Note: image_dataset_from_directory handles batching, so no need to set batch_size here.
    epochs=250,
    callbacks=[checkpoint_cb, early_stopping_cb]
)
```

```
1/3 ━━━━━━━━━━━━━━━━━━━━ 0s 74ms/step - accuracy: 1.0000 - loss: 0.0042
Epoch 25: val_accuracy did not improve from 0.88889
3/3 ━━━━━━━━━━━━━━━━━━━━ 0s 38ms/step - accuracy: 1.0000 - loss: 0.0037 - val_accuracy: 0.7778 - val_loss: 0.4866
Epoch 26/250
1/3 ━━━━━━━━━━━━━━━━━━━━ 0s 45ms/step - accuracy: 1.0000 - loss: 0.0028
Epoch 26: val_accuracy did not improve from 0.88889
3/3 ━━━━━━━━━━━━━━━━━━━━ 0s 38ms/step - accuracy: 1.0000 - loss: 0.0044 - val_accuracy: 0.7778 - val_loss: 0.4970
Epoch 26: early stopping
Restoring model weights from the end of the best epoch: 16.
```

Task 5: Evaluate the Model

```
# Define path to the test dataset folder
test_dir = "/content/FruitinAmazon/test"

# Load test dataset
test_dataset = tf.keras.preprocessing.image_dataset_from_directory(
    test_dir,
    labels='inferred',
    label_mode='int',
    image_size=(img_height, img_width),
    interpolation='nearest',
    batch_size=batch_size,
    shuffle=False
)

# Apply normalization to test dataset (same as for train and val)
test_dataset = test_dataset.map(lambda x, y: (rescale(x), y))

# Now, evaluate the model
test_loss, test_accuracy = model.evaluate(test_dataset, verbose=1)

print(f"Test Accuracy: {test_accuracy:.4f}")
print(f"Test Loss: {test_loss:.4f}")
```

```
Found 30 files belonging to 6 classes.
1/1 ━━━━━━━━━━━━━━━━━━━━ 1s 824ms/step - accuracy: 0.5333 - loss: 1.0288
Test Accuracy: 0.5333
Test Loss: 1.0288
```

Task 6: Save and Load the Model

```
import tensorflow as tf
from tensorflow.keras.models import load_model

# Step 1: Save the trained model
model.save("my_cnn_model.h5")
print("Model saved successfully!")

# Step 2: Load the saved model
loaded_model = load_model("my_cnn_model.h5")
print("Model loaded successfully!")

# Step 3: Re-evaluate the model on the test dataset
# Ensure test_dataset is defined before evaluating
test_loss, test_accuracy = loaded_model.evaluate(test_dataset, verbose=1)

print(f"Test Accuracy: {test_accuracy:.4f}")
print(f"Test Loss: {test_loss:.4f}")
```

```
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is considered legacy. We recommend using instead the native Keras format, e.g. `model.save('my_model.keras')` or `keras.saving.save_model(model, 'my_model.keras')`.
WARNING:absl:Compiled the loaded model, but the compiled metrics have yet to be built. `model.compile_metrics` will be empty until you train or evaluate the model.
Model saved successfully!
Model loaded successfully!
1/1 ━━━━━━━━━━━━━━━━━━━━ 1s 549ms/step - accuracy: 0.5333 - loss: 1.0288
Test Accuracy: 0.5333
Test Loss: 1.0288
```

Task 7: Predictions and Classification Report

```
# Make predictions
y_pred_probs = loaded_model.predict(test_dataset)  # Use test_dataset instead of X_test
y_pred = np.argmax(y_pred_probs, axis=1)

# Get true labels from test_dataset
y_true = []
for images, labels in test_dataset:
    y_true.extend(labels.numpy())

# Classification report
!pip install scikit-learn
from sklearn.metrics import classification_report
print("Classification Report:")
print(classification_report(y_true, y_pred))  # Use y_true instead of y_test

# Plot training & validation accuracy and loss
import matplotlib.pyplot as plt
plt.figure(figsize=(12, 5))

# Accuracy plot
plt.subplot(1, 2, 1)
plt.plot(history.history['accuracy'], label='Train Accuracy')
plt.plot(history.history['val_accuracy'], label='Validation Accuracy')
plt.legend()
plt.title("Accuracy")

# Loss plot
plt.subplot(1, 2, 2)
plt.plot(history.history['loss'], label='Train Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.legend()
plt.title("Loss")

plt.show()
```

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 1s 508ms/step
Requirement already satisfied: scikit-learn in /usr/local/lib/python3.11/dist-packages (1.6.1)
Requirement already satisfied: numpy>=1.19.5 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (2.0.2)
Requirement already satisfied: scipy>=1.6.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.14.1)
Requirement already satisfied: joblib>=1.2.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (1.4.2)
Requirement already satisfied: threadpoolctl>=3.1.0 in /usr/local/lib/python3.11/dist-packages (from scikit-learn) (3.6.0)
Classification Report:
              precision    recall  f1-score   support

           0       0.75      0.60      0.67         5
           1       0.38      0.60      0.46         5
           2       0.75      0.60      0.67         5
           3       0.50      0.80      0.62         5
           4       0.50      0.20      0.29         5
           5       0.50      0.40      0.44         5

    accuracy                           0.53        30
   macro avg       0.56      0.53      0.52        30
weighted avg       0.56      0.53      0.52        30
```