

## ✓ Task 1: Data Understanding and Visualization:

```
from google.colab import drive
drive.mount('/content/drive')
```

↻ Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force\_remount=True).

- ✓ 1. Load and visualize images from a dataset stored in directories, where each subdirectory represents a class.

```
import os
import numpy as np
import tensorflow as tf
from tensorflow.keras.utils import to_categorical
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
from PIL import Image

# Training and testing directory
train_dir = "/content/drive/MyDrive/AI and Machine Learning/Workshop5/train"
test_dir = "/content/drive/MyDrive/AI and Machine Learning/Workshop5/test"

img_height, img_width = 128, 128 # Increased resolution

def load_images_from_directory(directory):
    images = []
    labels = []
    class_names = sorted(os.listdir(directory)) # Ensure consistent label order
    class_dict = {class_name: idx for idx, class_name in enumerate(class_names)}

    for class_name in class_names:
        class_path = os.path.join(directory, class_name)
        if not os.path.isdir(class_path):
            continue

        for img_name in os.listdir(class_path):
            img_path = os.path.join(class_path, img_name)
            try:
                img = Image.open(img_path)
                img = img.resize((img_width, img_height), Image.LANCZOS) # LANCZOS for sharper resizing
                images.append(np.array(img))
                labels.append(class_dict[class_name])
            except Exception as e:
                print(f"Error loading image {img_path}: {e}")

    return np.array(images), np.array(labels), class_names

# Load training images
X, y, class_names = load_images_from_directory(train_dir)

# Normalize pixel values to [0,1]
X = X / 255.0

# Convert labels to categorical
y = to_categorical(y, num_classes=len(class_names))

# Split data into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)

# Display some sample images
def display_sample_images(X, y, class_names, rows=2, cols=5):
    fig, axes = plt.subplots(rows, cols, figsize=(10, 5))
    axes = axes.flatten()

    for i in range(rows * cols):
        idx = np.random.randint(len(X))
        axes[i].imshow(X[idx], interpolation='nearest') # Ensure sharp display
        axes[i].set_title(class_names[np.argmax(y[idx])])
        axes[i].axis('off')
```

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

# Display sample images from training set
display_sample_images(X_train, y_train, class_names)
```



## 2. Check for Corrupted Image:

```
import os
from PIL import Image

# Training directory
train_dir = "/content/drive/MyDrive/AI and Machine Learning/Workshop5/train"

def remove_corrupted_images(directory):
    corrupted_images = []

    # Iterate through each class subdirectory
    for class_name in os.listdir(directory):
        class_path = os.path.join(directory, class_name)
        if not os.path.isdir(class_path):
            continue

        # Iterate through each image in the class subdirectory
        for img_name in os.listdir(class_path):
            img_path = os.path.join(class_path, img_name)

            try:
                # Attempt to open the image
                img = Image.open(img_path)
                img.verify() # Verify the image is valid
            except (IOError, SyntaxError) as e:
                # If an error occurs, it's a corrupted image
                corrupted_images.append(img_path)
                os.remove(img_path) # Remove corrupted image
                print(f"Removed corrupted image: {img_path}")

    # Report if no corrupted images were found
    if not corrupted_images:
        print("No corrupted images found.")

# Call the function to check and remove corrupted images
remove_corrupted_images(train_dir)
```

No corrupted images found.

## ✓ Task 2: Loading and Preprocessing Image Data in keras:

```
# Define image size and batch size
img_height = 128
img_width = 128
batch_size = 32
validation_split=0.2 #80% training , 20% validation
# Create preprocessing layer for normalization
rescale = tf.keras.layers.Rescaling(1./255) # Normalize pixel values to [0,1]

train_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=True,
    validation_split=validation_split,
    subset='training',
    seed=123
)

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

# Create validation dataset with normalization
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,
    validation_split=validation_split,
    subset='validation',
    seed=123
)
# Apply the normalization (Rescaling) to the validation dataset
val_ds = val_ds.map(lambda x, y: (rescale(x), y))
```

➡ Found 90 files belonging to 6 classes.  
Using 72 files for training.  
Found 90 files belonging to 6 classes.  
Using 18 files for validation.

## ✓ Task 3 - Implement a CNN with

```
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense
from tensorflow.keras.optimizers import Adam

# Define the CNN + Fully Connected Network model
model = Sequential()

# Convolutional Layer 1
model.add(Conv2D(32, (3, 3), padding='same', strides=1, activation='relu', input_shape=(128, 128, 3)))

# Max Pooling Layer 1
model.add(MaxPooling2D(pool_size=(2, 2), strides=2))

# Convolutional Layer 2
model.add(Conv2D(32, (3, 3), padding='same', strides=1, activation='relu'))

# Max Pooling Layer 2
model.add(MaxPooling2D(pool_size=(2, 2), strides=2))

# Flatten the output from the convolutional layers
model.add(Flatten())

# Hidden Layer 1 - 64 neurons
```

```

model.add(Dense(64, activation='relu'))

# Hidden Layer 2 - 128 neurons
model.add(Dense(128, activation='relu'))

# Output Layer (Number of classes = len(class_names))
model.add(Dense(len(class_names), activation='softmax'))

# Compile the model
model.compile(optimizer=Adam(), loss='categorical_crossentropy', metrics=['accuracy'])

# 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_3 (Dense)	(None, 64)	2,097,216
dense_4 (Dense)	(None, 128)	8,320
dense_5 (Dense)	(None, 6)	774

Total params: 2,116,454 (8.07 MB)  
 Trainable params: 2,116,454 (8.07 MB)  
 Non-trainable params: 0 (0.00 B)

Explanation of the Layers: Convolutional Layers (Conv2D) and Max Pooling Layers (MaxPooling2D): These layers are the same as in the previous CNN model. They extract features from the image and reduce spatial dimensions.

Flatten Layer:

The Flatten() layer reshapes the output from the convolutional layers into a 1D vector that can be passed to the fully connected layers.

Hidden Layers:

Dense Layer 1: Has 64 neurons, with ReLU activation. This layer learns the relationships between the features extracted by the convolutional layers.

Dense Layer 2: Has 128 neurons, also with ReLU activation. This further processes the features learned in the first hidden layer.

Output Layer:

The number of neurons is equal to the number of classes (i.e., len(class\_names)).

Softmax activation is used for multi-class classification, where the model outputs probabilities for each class.

Model Compilation: Optimizer: Adam optimizer is used for gradient descent.

Loss function: categorical\_crossentropy is used for multi-class classification.

Metrics: Accuracy is used to evaluate the model's performance.

## ✓ Task 4: Compile the Model

```

# Compile the model
model.compile(
    optimizer='adam', # Adam optimizer
    loss='sparse_categorical_crossentropy', # Use 'categorical_crossentropy' if labels are one-hot encoded
    metrics=['accuracy'] # Accuracy metric
)

```

## ✓ Task 4: Train the Model

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

```
# Model Compilation using sparse_categorical_crossentropy
model.compile(
    optimizer='adam', # Adam optimizer
    loss='sparse_categorical_crossentropy', # For integer labels
    metrics=['accuracy'] # Accuracy metric
)
```

```
# Define callbacks
checkpoint_callback = ModelCheckpoint(
    'best_model.h5',
    monitor='val_loss',
    save_best_only=True,
    mode='min',
    verbose=1
)
```

```
# Define callbacks
checkpoint_callback = ModelCheckpoint(
    'best_model.h5',
    monitor='val_loss',
    save_best_only=True,
    mode='min',
    verbose=1
)
```

```
# Train the model using model.fit() with callbacks
history = model.fit(
    X_train, # Training data
    y_train, # Training labels
    epochs=250,
    batch_size=16,
    validation_data=(X_val, y_val),
    callbacks=[checkpoint_callback, early_stopping_callback]
)
```

```

WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is con
5/5 ----- 3s 357ms/step - accuracy: 0.9094 - loss: 0.3408 - val_accuracy: 0.6111 - val_loss: 1.6419
Epoch 2/250
5/5 ----- 0s 256ms/step - accuracy: 0.8684 - loss: 0.4122
Epoch 2: val_loss did not improve from 1.64195
5/5 ----- 2s 302ms/step - accuracy: 0.8649 - loss: 0.4222 - val_accuracy: 0.3889 - val_loss: 2.0205
Epoch 3/250
5/5 ----- 0s 250ms/step - accuracy: 0.9066 - loss: 0.3396
Epoch 3: val_loss improved from 1.64195 to 1.50525, saving model to best_model.h5
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is con
5/5 ----- 3s 319ms/step - accuracy: 0.9060 - loss: 0.3349 - val_accuracy: 0.5556 - val_loss: 1.5052
Epoch 4/250
5/5 ----- 0s 424ms/step - accuracy: 0.9611 - loss: 0.2082
Epoch 4: val_loss improved from 1.50525 to 1.24996, saving model to best_model.h5
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is con
5/5 ----- 4s 517ms/step - accuracy: 0.9630 - loss: 0.2056 - val_accuracy: 0.5556 - val_loss: 1.2500
Epoch 5/250
5/5 ----- 0s 261ms/step - accuracy: 0.9837 - loss: 0.0833
Epoch 5: val_loss did not improve from 1.24996
5/5 ----- 4s 307ms/step - accuracy: 0.9841 - loss: 0.0813 - val_accuracy: 0.5000 - val_loss: 1.3805
Epoch 6/250
5/5 ----- 0s 253ms/step - accuracy: 1.0000 - loss: 0.0572
Epoch 6: val_loss did not improve from 1.24996
5/5 ----- 2s 288ms/step - accuracy: 1.0000 - loss: 0.0562 - val_accuracy: 0.5000 - val_loss: 1.3517
Epoch 7/250
5/5 ----- 0s 257ms/step - accuracy: 1.0000 - loss: 0.0268
Epoch 7: val_loss did not improve from 1.24996
5/5 ----- 3s 303ms/step - accuracy: 1.0000 - loss: 0.0264 - val_accuracy: 0.5556 - val_loss: 1.2654
Epoch 8/250
5/5 ----- 0s 437ms/step - accuracy: 1.0000 - loss: 0.0124
Epoch 8: val_loss did not improve from 1.24996
5/5 ----- 4s 529ms/step - accuracy: 1.0000 - loss: 0.0123 - val_accuracy: 0.5556 - val_loss: 1.4372
Epoch 9/250
5/5 ----- 0s 419ms/step - accuracy: 1.0000 - loss: 0.0146

```

```

5/5 ----- 0s 240ms/step - accuracy: 1.0000 - loss: 0.0048
Epoch 10: val_loss did not improve from 1.24996
5/5 ----- 2s 281ms/step - accuracy: 1.0000 - loss: 0.0048 - val_accuracy: 0.6111 - val_loss: 1.4484
Epoch 11/250
5/5 ----- 0s 252ms/step - accuracy: 1.0000 - loss: 0.0044
Epoch 11: val_loss did not improve from 1.24996
5/5 ----- 1s 291ms/step - accuracy: 1.0000 - loss: 0.0044 - val_accuracy: 0.5556 - val_loss: 1.4852
Epoch 12/250
5/5 ----- 0s 257ms/step - accuracy: 1.0000 - loss: 0.0033
Epoch 12: val_loss did not improve from 1.24996
5/5 ----- 1s 296ms/step - accuracy: 1.0000 - loss: 0.0032 - val_accuracy: 0.5000 - val_loss: 1.5288
Epoch 13/250
5/5 ----- 0s 252ms/step - accuracy: 1.0000 - loss: 0.0021
Epoch 13: val_loss did not improve from 1.24996
5/5 ----- 2s 299ms/step - accuracy: 1.0000 - loss: 0.0021 - val_accuracy: 0.5000 - val_loss: 1.5915
Epoch 14/250
5/5 ----- 0s 260ms/step - accuracy: 1.0000 - loss: 0.0016
Epoch 14: val_loss did not improve from 1.24996
5/5 ----- 3s 306ms/step - accuracy: 1.0000 - loss: 0.0016 - val_accuracy: 0.5000 - val_loss: 1.6524
Epoch 14: early stopping

```

```

# Remove one-hot encoding (to_categorical)
X, y, class_names = load_images_from_directory(train_dir)

# Normalize pixel values to [0,1]
X = X / 255.0

# Split data into training and validation sets
X_train, X_val, y_train, y_val = train_test_split(X, y, test_size=0.2, random_state=42)

# Model Compilation using sparse_categorical_crossentropy
model.compile(
    optimizer='adam', # Adam optimizer
    loss='sparse_categorical_crossentropy', # For integer labels
    metrics=['accuracy'] # Accuracy metric
)

# Define callbacks
checkpoint_callback = ModelCheckpoint(
    'best_model.h5', # File path to save the best model
    monitor='val_loss', # Monitor validation loss
    save_best_only=True, # Save only the best model
    mode='min', # Minimize the validation loss
    verbose=1 # Print a message when the model is saved
)

early_stopping_callback = EarlyStopping(
    monitor='val_loss', # Monitor validation loss
    patience=10, # Stop after 10 epochs with no improvement
    restore_best_weights=True, # Restore the weights of the best model
    verbose=1 # Print a message when training stops
)

# Train the model using model.fit() with callbacks
history = model.fit(
    X_train, # Training data
    y_train, # Training labels
    epochs=250, # Number of epochs
    batch_size=16, # Batch size
    validation_data=(X_val, y_val), # Validation data
    callbacks=[checkpoint_callback, early_stopping_callback] # Callbacks for saving the best model and early stopping
)

5/5 ----- 4s 375ms/step - accuracy: 0.9476 - loss: 0.1434 - val_accuracy: 0.6111 - val_loss: 2.1825
Epoch 2/250
5/5 ----- 0s 251ms/step - accuracy: 0.9941 - loss: 0.0842
Epoch 2: val_loss improved from 2.18252 to 1.59316, saving model to best_model.h5
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is con
5/5 ----- 2s 315ms/step - accuracy: 0.9928 - loss: 0.0853 - val_accuracy: 0.5556 - val_loss: 1.5932
Epoch 3/250
5/5 ----- 0s 408ms/step - accuracy: 0.9712 - loss: 0.1023
Epoch 3: val_loss improved from 1.59316 to 1.41539, saving model to best_model.h5
WARNING:absl:You are saving your model as an HDF5 file via `model.save()` or `keras.saving.save_model(model)`. This file format is con
5/5 ----- 3s 503ms/step - accuracy: 0.9737 - loss: 0.0959 - val_accuracy: 0.5556 - val_loss: 1.4154

```

```

5/5 ----- 4s 321ms/step - accuracy: 0.9893 - loss: 0.0775 - val_accuracy: 0.5556 - val_loss: 1.3472
Epoch 5/250
5/5 ----- 0s 241ms/step - accuracy: 1.0000 - loss: 0.0156
Epoch 5: val_loss did not improve from 1.34722
5/5 ----- 2s 287ms/step - accuracy: 1.0000 - loss: 0.0157 - val_accuracy: 0.5556 - val_loss: 1.6699
Epoch 6/250
5/5 ----- 0s 241ms/step - accuracy: 0.9899 - loss: 0.0249
Epoch 6: val_loss did not improve from 1.34722
5/5 ----- 2s 276ms/step - accuracy: 0.9893 - loss: 0.0248 - val_accuracy: 0.5556 - val_loss: 1.7028
Epoch 7/250
5/5 ----- 0s 245ms/step - accuracy: 1.0000 - loss: 0.0038
Epoch 7: val_loss did not improve from 1.34722
5/5 ----- 1s 284ms/step - accuracy: 1.0000 - loss: 0.0038 - val_accuracy: 0.6667 - val_loss: 1.7222
Epoch 8/250
5/5 ----- 0s 424ms/step - accuracy: 1.0000 - loss: 0.0034
Epoch 8: val_loss did not improve from 1.34722
5/5 ----- 2s 516ms/step - accuracy: 1.0000 - loss: 0.0036 - val_accuracy: 0.6667 - val_loss: 1.9073
Epoch 9/250
5/5 ----- 0s 387ms/step - accuracy: 1.0000 - loss: 0.0040
Epoch 9: val_loss did not improve from 1.34722
5/5 ----- 2s 438ms/step - accuracy: 1.0000 - loss: 0.0039 - val_accuracy: 0.6111 - val_loss: 1.9817
Epoch 10/250
5/5 ----- 0s 240ms/step - accuracy: 1.0000 - loss: 0.0025
Epoch 10: val_loss did not improve from 1.34722
5/5 ----- 1s 287ms/step - accuracy: 1.0000 - loss: 0.0024 - val_accuracy: 0.6111 - val_loss: 2.0376
Epoch 11/250
5/5 ----- 0s 248ms/step - accuracy: 1.0000 - loss: 9.8660e-04
Epoch 11: val_loss did not improve from 1.34722
5/5 ----- 1s 297ms/step - accuracy: 1.0000 - loss: 9.7746e-04 - val_accuracy: 0.5556 - val_loss: 2.0715
Epoch 12/250
5/5 ----- 0s 249ms/step - accuracy: 1.0000 - loss: 7.4941e-04
Epoch 12: val_loss did not improve from 1.34722
5/5 ----- 3s 296ms/step - accuracy: 1.0000 - loss: 7.3726e-04 - val_accuracy: 0.5000 - val_loss: 2.0307
Epoch 13/250
5/5 ----- 0s 247ms/step - accuracy: 1.0000 - loss: 4.4916e-04
Epoch 13: val_loss did not improve from 1.34722
5/5 ----- 1s 293ms/step - accuracy: 1.0000 - loss: 4.5313e-04 - val_accuracy: 0.5000 - val_loss: 1.9863
Epoch 14/250
5/5 ----- 0s 247ms/step - accuracy: 1.0000 - loss: 4.1652e-04
Epoch 14: val_loss did not improve from 1.34722
5/5 ----- 3s 294ms/step - accuracy: 1.0000 - loss: 4.1282e-04 - val_accuracy: 0.5000 - val_loss: 1.9597
Epoch 14: early stopping

```

## ▼ Task 5: Evaluate the Model

```

from tensorflow.keras.preprocessing import image_dataset_from_directory

# Load the test data (assuming the test data is in a similar format to the training data)
test_ds = image_dataset_from_directory(
    test_dir,
    labels='inferred',
    label_mode='int',
    image_size=(img_height, img_width), # Ensure test images are resized to match training images
    interpolation='nearest',
    batch_size=batch_size,
    shuffle=False
)

# Apply normalization to the test dataset (same as training and validation datasets)
test_ds = test_ds.map(lambda x, y: (rescale(x), y))

# Evaluate the model on the test dataset
test_loss, test_accuracy = model.evaluate(test_ds)

# Print the results
print(f"Test Loss: {test_loss}")
print(f"Test Accuracy: {test_accuracy}")

Found 30 files belonging to 6 classes.
1/1 ----- 3s 3s/step - accuracy: 0.7000 - loss: 1.0394
Test Loss: 1.0393983125686646
Test Accuracy: 0.699999988079071

# Save the model to an .h5 file
model.save('my_model.keras')

```

```

from tensorflow.keras.models import load_model

# Load the model in the Keras format
loaded_model = load_model('my_model.keras')

↗ /usr/local/lib/python3.11/dist-packages/keras/src/saving/saving_lib.py:757: UserWarning: Skipping variable loading for optimizer 'rmsprop'
saveable.load_own_variables(weights_store.get(inner_path))

# Evaluate the loaded model on the test dataset
test_loss, test_accuracy = loaded_model.evaluate(test_ds)

# Print the results
print(f"Test Loss (after reloading): {test_loss}")
print(f"Test Accuracy (after reloading): {test_accuracy}")

↗ 1/1 ————— 1s 572ms/step - accuracy: 0.7000 - loss: 1.0394
Test Loss (after reloading): 1.0393983125686646
Test Accuracy (after reloading): 0.699999988079071

```

## ✓ Task 7: Predictions and Classification Report

```

import numpy as np
from sklearn.metrics import classification_report
import tensorflow as tf
import os

# Get class names from the directory structure
class_names = sorted(os.listdir(test_dir)) # List of class names

# Get the test dataset (make sure it's in the same format as train_ds)
test_ds = tf.keras.preprocessing.image_dataset_from_directory(
    test_dir,
    labels='inferred',
    label_mode='int',
    image_size=(img_height, img_width),
    batch_size=batch_size,
    shuffle=False
)

# Get true labels from the test dataset
true_labels = np.concatenate([y.numpy() for _, y in test_ds], axis=0)

# Make predictions on the test dataset
predictions = loaded_model.predict(test_ds)

# Convert predicted probabilities to class labels
predicted_labels = np.argmax(predictions, axis=-1)

# Ensure true_labels and predicted_labels are 1D arrays
true_labels = true_labels.flatten()
predicted_labels = predicted_labels.flatten()

# Generate the classification report
report = classification_report(true_labels, predicted_labels, target_names=class_names)

# Print the classification report
print(report)

↗ Found 30 files belonging to 6 classes.
1/1 ————— 0s 370ms/step

```

	precision	recall	f1-score	support
acai	0.75	0.60	0.67	5
cupuacu	0.38	0.60	0.46	5
graviola	0.71	1.00	0.83	5
guarana	0.83	1.00	0.91	5
pupunha	1.00	0.60	0.75	5
tucuma	1.00	0.40	0.57	5
accuracy			0.70	30
macro avg	0.78	0.70	0.70	30
weighted avg	0.78	0.70	0.70	30



Start coding or [generate](#) with AI.