

Data Collection and Preprocessing Phase

Date	16 JUNE 2025
Team ID	SWTID1749825524
Project Title	Deepfruitveg: Automated Fruit and Veg Identification
Maximum Marks	6 Marks

Data Exploration and Preprocessing Report

Dataset variables will be statistically analyzed to identify patterns and outliers, with Python employed for preprocessing tasks like normalization and feature engineering. Data cleaning will address missing values and outliers, ensuring quality for subsequent analysis and modeling, and forming a strong foundation for insights and predictions.

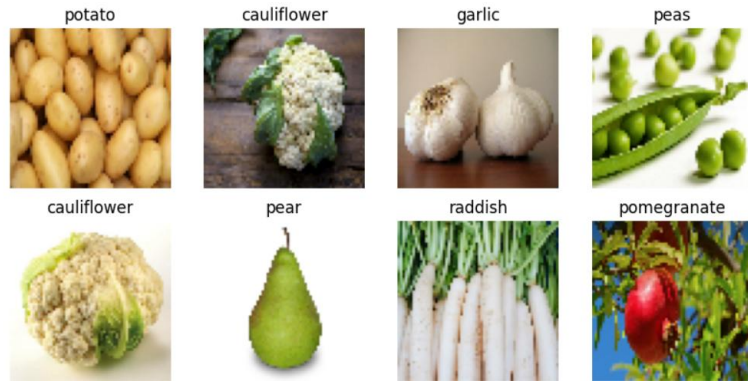
Section	Description
Data Overview	<p><u>Dimension:</u> TEST SET: Found 3125 files belonging to 36 classes.</p> <p>VALIDATION SET: Found 351 files belonging to 36 classes.</p> <p>TRAIN SET: Found 379 files belonging to 36 classes.</p>

Data overlook

```

In [ ]: def show_image_samples(dataset, class_names, num_samples=16):
        plt.figure(figsize=(10, 10))
        for images, labels in dataset.take(1): # Take one batch
            for i in range(min(num_samples, images.shape[0])):
                ax = plt.subplot(4, 4, i + 1)
                plt.imshow(images[i].numpy().astype("uint8"))
                predicted_class_index = np.argmax(labels[i])
                plt.title(class_names[predicted_class_index])
                plt.axis("off")
            plt.show()

        show_image_samples(test_set, test_set.class_names)
  
```



Data Preprocessing Code Screenshots

Loading Data

```
importing dataset

In [ ]:
from google.colab import drive
drive.mount('/content/drive')

Mounted at /content/drive

In [ ]:
import os
print(os.listdir('/content/drive/MyDrive'))

In [ ]:
print(os.listdir('/content/drive/MyDrive/ai'))

['validation', 'train', 'test']
```

Importing Libraries

```
importing libraries

In [ ]:
import numpy as np
import tensorflow as tf
import matplotlib.pyplot as plt
```

Training image preprocessing

```
In [ ]:
training_set = tf.keras.utils.image_dataset_from_directory(
    '/content/drive/MyDrive/ai/test',
    labels='inferred',
    label_mode='categorical',
    class_names=None,
    color_mode='rgb',
    batch_size=32,
    image_size=(64, 64),
    shuffle=True,
    seed=None,
    validation_split=None,
    subset=None,
    interpolation='bilinear',
    follow_links=False,
    crop_to_aspect_ratio=False,
)

Found 3125 files belonging to 36 classes.
```

Validation image preprocessing

validation image preprocessing

In []:

```
validation_set = tf.keras.utils.image_dataset_from_directory(  
    '/content/drive/MyDrive/ai/validation',  
    labels="inferred",  
    label_mode="categorical",  
    class_names=None,  
    color_mode="rgb",  
    batch_size=32,  
    image_size=(64, 64),  
    shuffle=True,  
    seed=None,  
    validation_split=None,  
    subset=None,  
    interpolation="bilinear",  
    follow_links=False,  
    crop_to_aspect_ratio=False  
)
```

Found 351 files belonging to 36 classes.