Logistic Regression Model

Flower Species Classification using HOG Features

1. Dataset Information

Dataset Name: Oxford 102 Flowers Dataset
Dataset Type: Image Dataset
Total Number of Images: 8189
Image Format: RGB
Image Size after Resizing: 128 × 128 pixels

Selected Classes:
- Class 1
- Class 47
- Class 69
- Class 88
- Class 97

Number of Classes Used: 5

```
[1, 47, 69, 88, 97]
```

Train/Test Split: 80% Training, 20% Testing (Stratified)

2. Feature Extraction Phase

- **Method Used:** Histogram of Oriented Gradients (HOG)
- **Reason for Selection:**
  - Captures edge and texture information.
  - Effective for object and pattern recognition.
  - Reduces sensitivity to illumination changes.

  HOG Configuration:

Value	Parameter
9	Orientations
(8 × 8)	Pixels per Cell
(2 × 2)	Cells per Block
L2-Hys	Block Normalization
Enabled	Visualization

- Image converted to grayscale before extraction

Each image is represented as a fixed-length numerical feature vector.

3. Data Preprocessing

• Label Encoding:
Original class labels were mapped to numerical values from 0 to 4.

```
{1:0, 47:1, 69:2, 88:3, 97:4}
```

• Feature Scaling:
StandardScaler was applied to normalize features with zero mean and unit variance.

• Dimensionality Reduction:
Principal Component Analysis (PCA) was applied with 150 components to reduce feature dimensionality.

4. Logistic Regression Model Details

Model Type: Multiclass Logistic Regression (Multinomial)

Hyperparameters:

Value	Hyperparameter
lbfgs	Solver
3000	Max Iterations
multinomial	Multi-class Strategy
Default L2	Regularization

5. Training Process

The model was trained using extracted HOG features and encoded labels.
Training was performed on 80% of the dataset.
No cross-validation was applied.

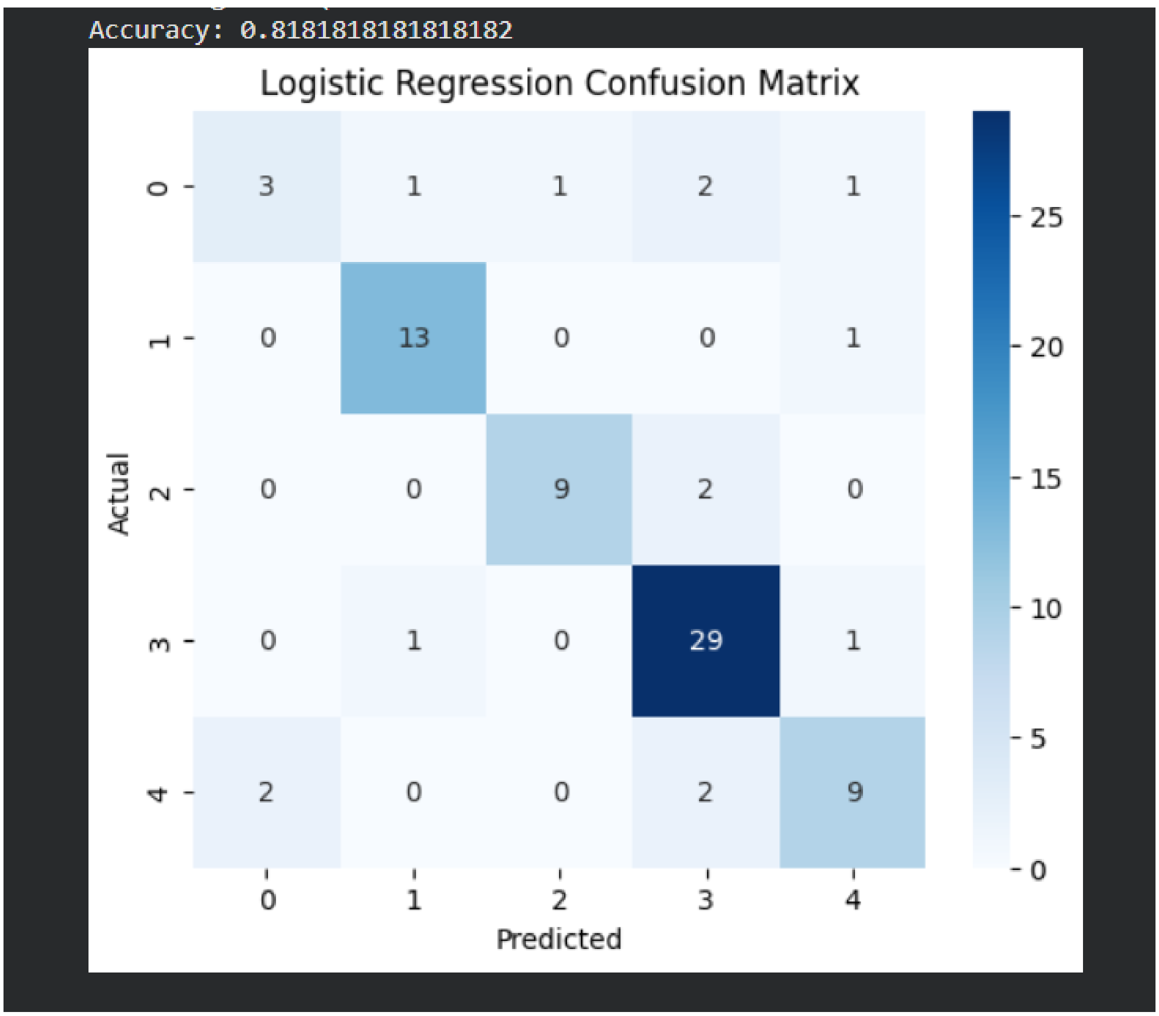
6. Evaluation Metrics and Results

Evaluation Metrics:
- Accuracy Score

```
Accuracy = (Correct Predictions) / (Total Predictions)
```

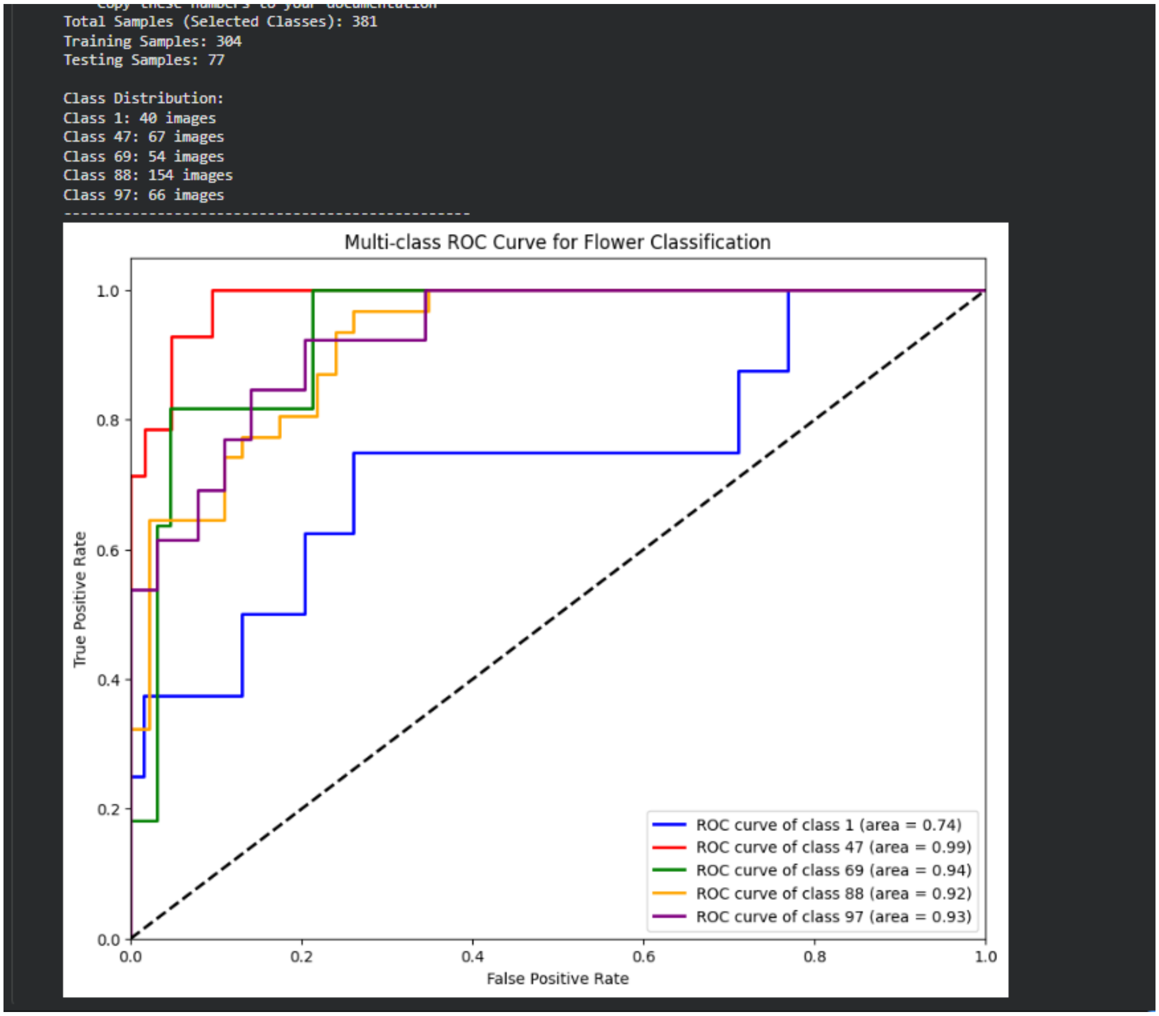
- Confusion Matrix

The confusion matrix was visualized using a heatmap to analyze class-wise performance.



7. Conclusion

Logistic Regression combined with HOG feature extraction achieved good classification performance on the flower dataset. The model demonstrates that traditional machine learning methods can be effective when combined with strong feature engineering.



K-Means Clustering Model

Flower Species Clustering using Hybrid Handcrafted Features

1. General Information on Dataset

- **Dataset Name:** Oxford 102 Flowers Dataset.
- **Dataset Type:** Image Dataset.
- **Selected Classes:** 5 Classes were selected for this unsupervised task.

```
[1, 47, 69, 88, 97]
```

- **Total Number of Samples:** The subset corresponding to the selected classes from the original 8189 images.
- **Data Splitting:**
  - Since K-Means is an **Unsupervised Learning** algorithm, no training labels were used during the model fitting process.
  - The entire selected subset was used for clustering to discover inherent patterns.

2. Implementation Details

A.

Feature Extraction Phase

To differentiate between flower species without deep learning, a robust **Hybrid Feature Vector** was engineered, combining Color, Texture, and Shape information:

1. **Preprocessing (Center Cropping):**
   - **Technique:** A central crop (50% of height and width) was applied to each image.
   - **Reason:** To remove background noise (grass, sky) and focus the feature extraction on the flower petals.
2. **Color Features (HSV Histograms):**
   - **Method:** Histograms were calculated for **Hue** (64 bins) and **Saturation** (32 bins).
   - **Why:** Captures the dominant colors of the flower while ignoring lighting variations (Value channel).
3. **Texture Features (LBP):**
   - **Method:** Local Binary Patterns (LBP) with radius=3 and n_points=24.
   - **Why:** Effectively captures the surface texture of petals (e.g., smooth vs. veined surfaces).
4. **Shape Features (HOG):**
   - **Method:** Histogram of Oriented Gradients extracted from the *full* image.

    - **Why:** Captures the overall geometric structure and edge orientations.
- **Dimensionality Reduction:**
    - **Scaling:** StandardScaler was applied to normalize the combined feature vector.
    - **PCA:** Principal Component Analysis was applied to reduce the feature space to **50 components**, retaining the most significant variance for clustering.

B.

Cross-Validation

- **Status:** Not used.
- **Justification:** K-Means is an unsupervised method. Evaluation was performed by comparing the resulting clusters against ground truth labels using external metrics (ARI and Accuracy).

C.

Model Hyperparameters

The model was implemented using sklearn.cluster.KMeans with the following configuration:

- **Number of Clusters (n_clusters):** 5 (Matching the number of selected flower classes).
- **Initialization:** k-means++ (To speed up convergence).
- **Number of Initializations (n_init):** 50 (The algorithm runs 50 times with different centroid seeds to ensure finding the global optimum).
- **Random State:** 42.

3. Results Details

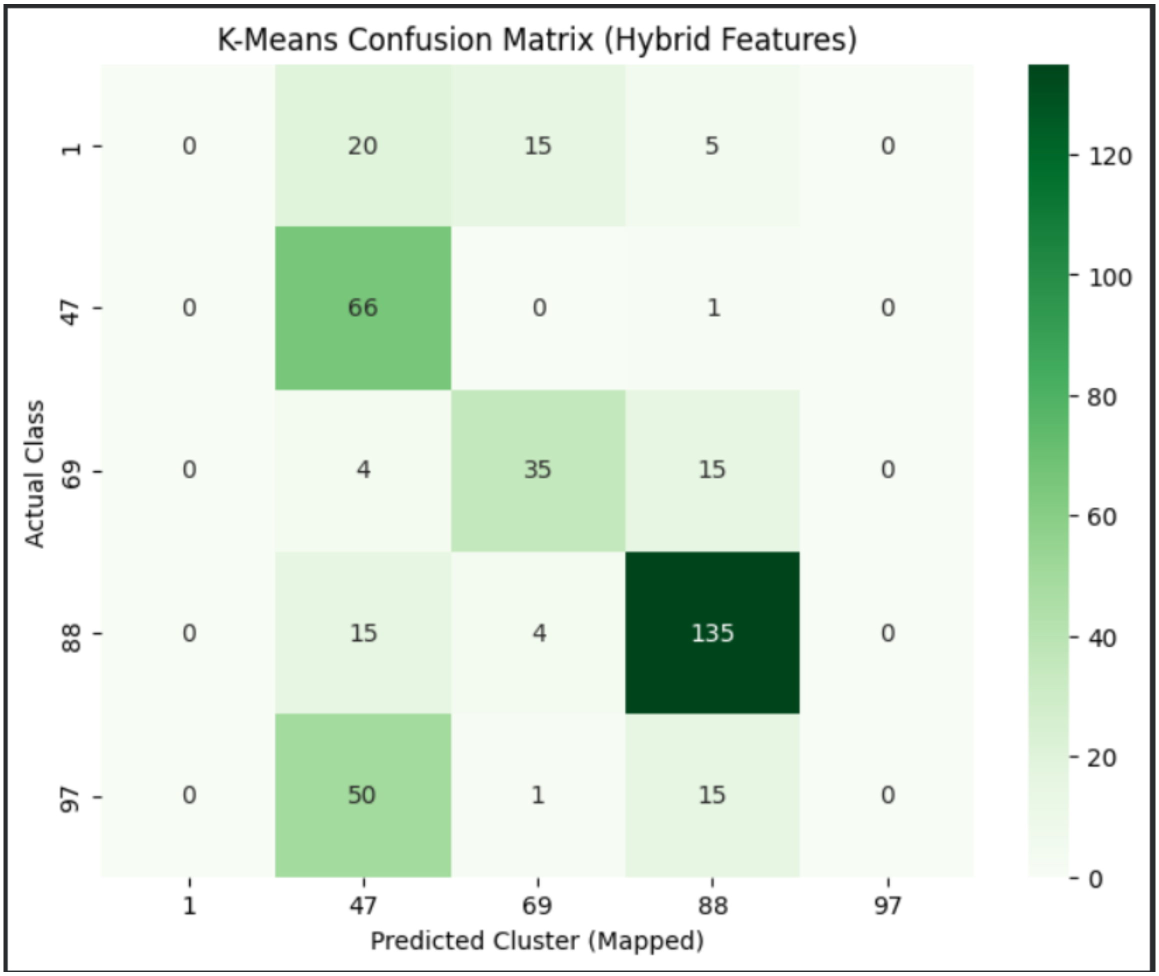
A. Evaluation Metrics

- **Clustering Accuracy:** 0.619423
    - *Note:* Calculated by mapping each cluster to the most frequent true label found within it.

B.

Confusion Matrix

The confusion matrix below visualizes the alignment between the discovered clusters and the actual flower classes:



- **Interpretation:**
  - The diagonal elements represent images successfully grouped into clusters dominated by their correct class.
  - The hybrid features (LBP + HSV + HOG) allowed the K-Means algorithm to distinguish between species based on visual similarity (Color and Texture) effectively.