Dogs Breeds Identification - DOCUMENTATION

1. Dataset Description:

The Stanford Dogs dataset contains images of 120 breeds of dogs. The dataset has been built using images and annotation from ImageNet for the task of fine-grained image categorization.

Number of categories: 120
Number of images: 20,580
Annotations: Class labels, Bounding boxes

2. Data Preprocessing Procedures:

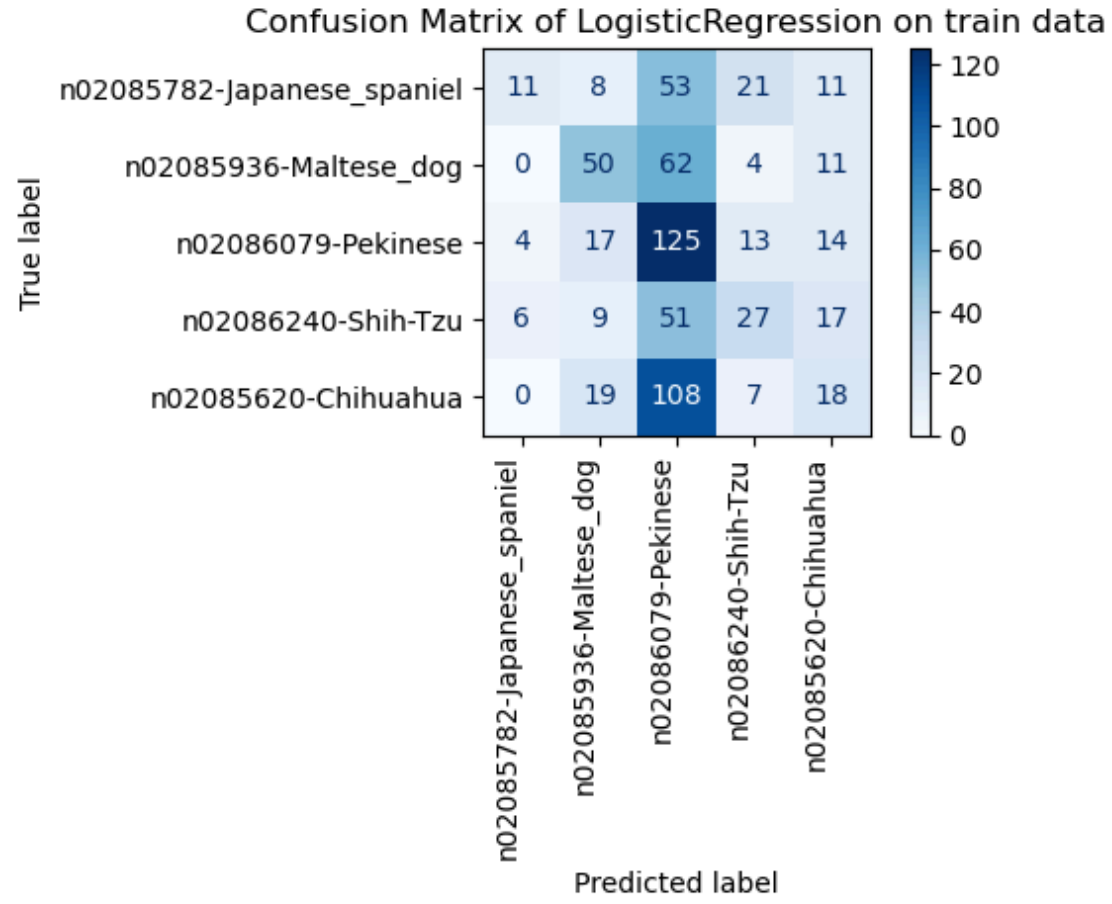
1. Resized the images from **(393 x 425)** to **(64 x 64)**.
2. Changed the colors from **RGB** to **LAB**.
3. Normalized the data by **255**.
4. PCA (Dimension Reduction), 92%. ***Before:*** *[952 × 8193]* ***After:*** *[952 × 142]*
5. Used label encoder to convert breeds from **object** and/or **string** to a numerical val.

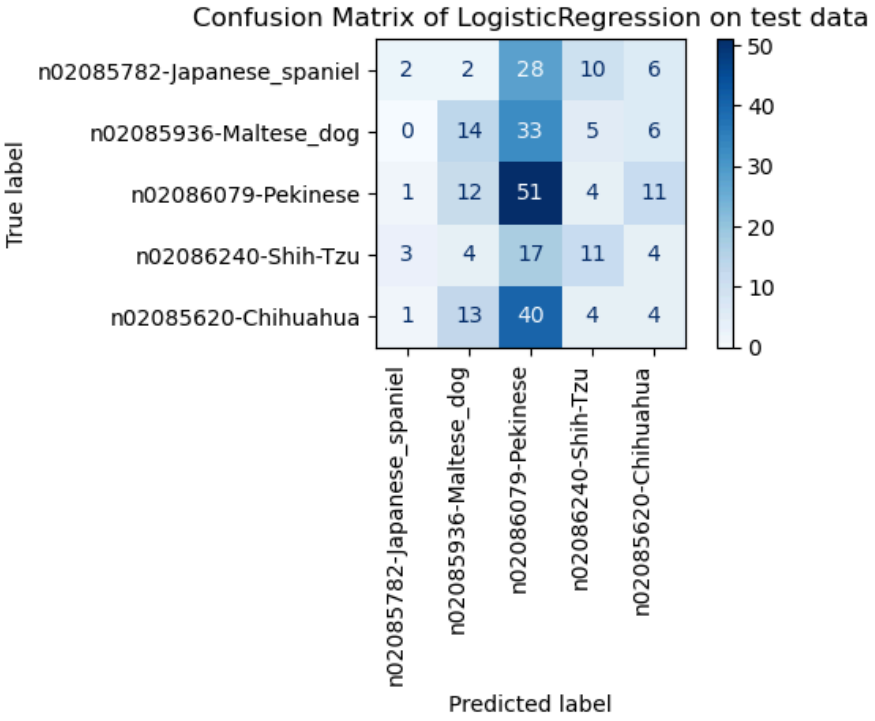
3. Shapes:

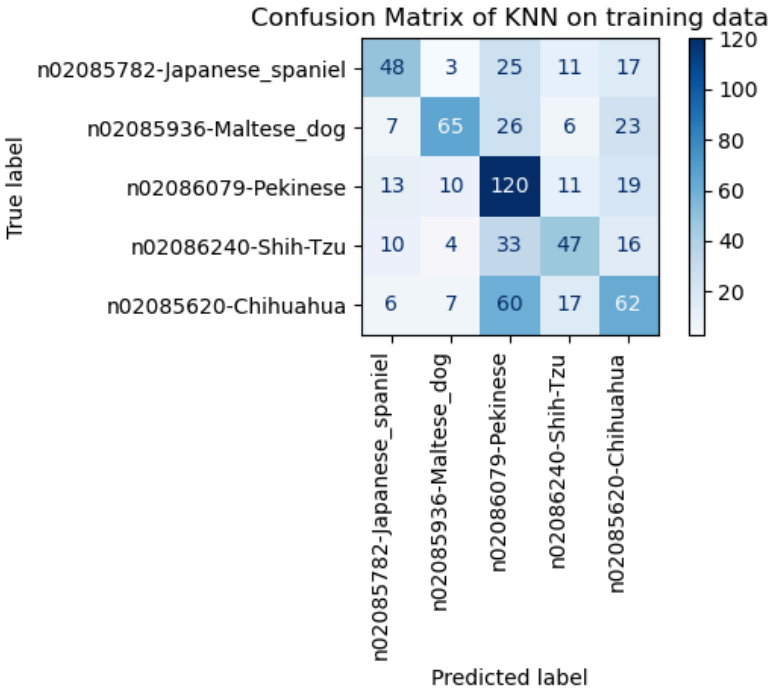
X-Train: (666, 142)
X-Test: (286, 142)
Y-Train: (666,)
Y-Test: (286,)

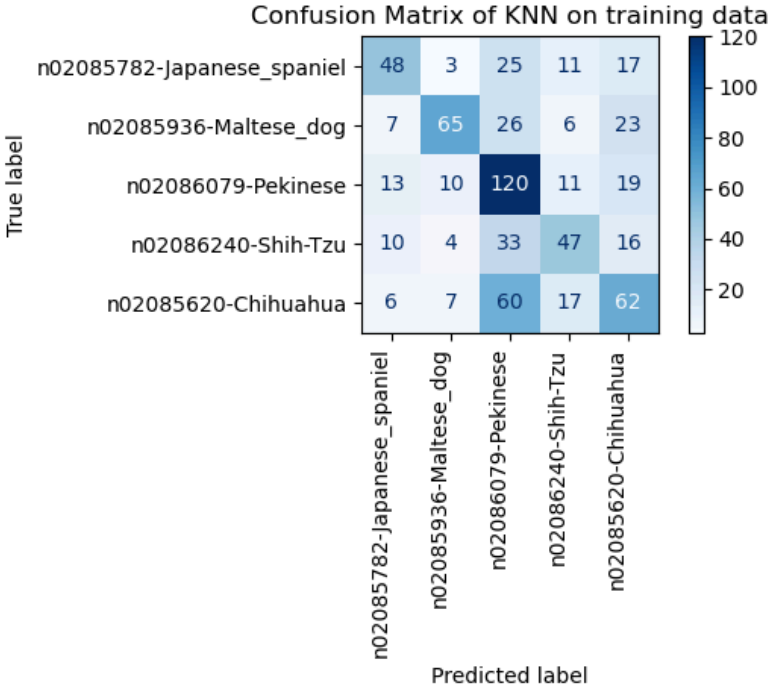
4. Used Algorithms:

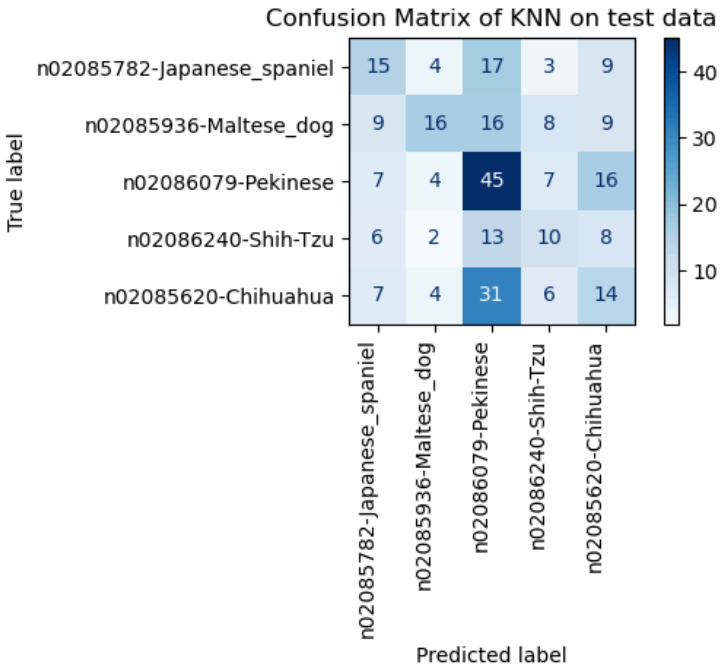
	Logistic Regression	KNN Classifier
Accuracy (Train Set)	34.68468468468468	51.35135135135135
Accuracy (Test Set)	28.671328671328673	34.96503496503497
ROC AUC Score (Train Set)	0.673467485560964	0.8425272808463472
ROC AUC Score (Test Set)	0.606422385822403	0.6060248879107233
Train Precision (Macro)	39.02101503588041	54.39470828737301
Test Precision (Macro)	27.02324429167226	35.74425060635283
Train Recall (Macro)	31.71777937716172	50.043171355449026
Test Recall (Macro)	25.503660166963975	32.80398060306467
Train Loss Curve	1.4869558149211097	0.9937528762900008
Test Loss Curve	1.5504039010213708	10.240998144909446

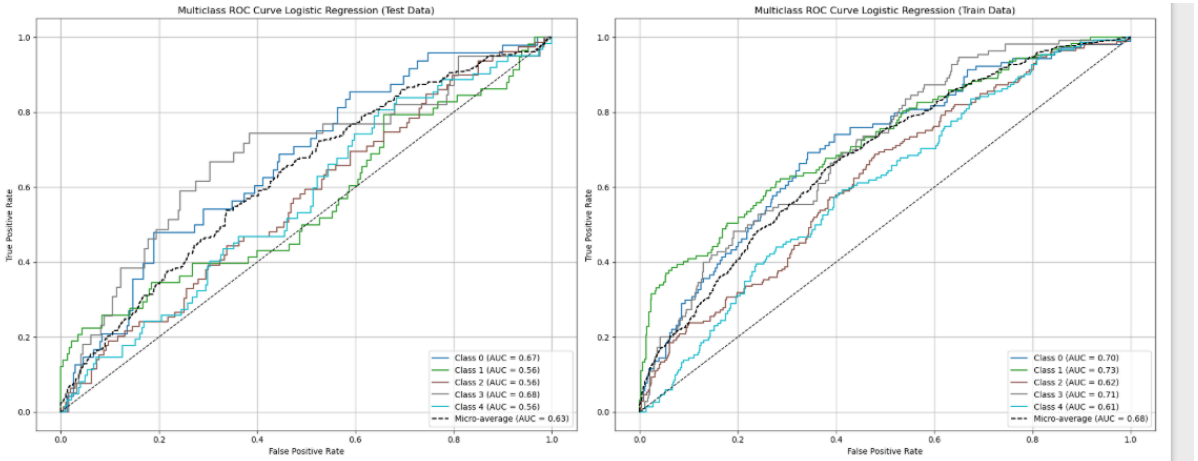

Confusion Matrix of LogisticRegression on train data
n02085782-Japanese_spaniel
n02085936-Maltese_dog
n02086079-Pekinese
n02086240-Shih-Tzu
n02085620-Chihuahua
11 8 53 21 11
0 50 62 4 11
4 17 125 13 14
6 9 51 27 17
0 19 108 7 18
120
100
80
60
40
20
0
True label
Predicted label

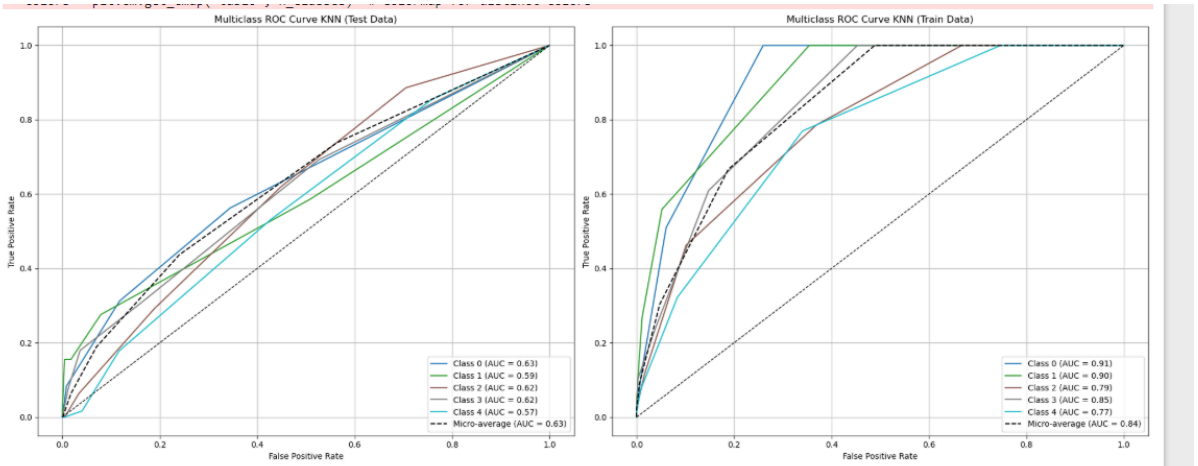

Confusion Matrix of LogisticRegression on test data
n02085782-Japanese_spaniel
n02085936-Maltese_dog
n02086079-Pekinese
n02086240-Shih-Tzu
n02085620-Chihuahua
2 2 28 10 6
0 14 33 5 6
1 12 51 4 11
3 4 17 11 4
1 13 40 4 4
50
40
30
20
10
0
True label
Predicted label


Confusion Matrix of KNN on training data
n02085782-Japanese_spaniel
n02085936-Maltese_dog
n02086079-Pekinese
n02086240-Shih-Tzu
n02085620-Chihuahua
True label
48 3 25 11 17
7 65 26 6 23
13 10 120 11 19
10 4 33 47 16
6 7 60 17 62
120
100
80
60
40
20
n02085782-Japanese_spaniel
n02085936-Maltese_dog
n02086079-Pekinese
n02086240-Shih-Tzu
n02085620-Chihuahua
Predicted label


Confusion Matrix of KNN on training data
n02085782-Japanese_spaniel
n02085936-Maltese_dog
n02086079-Pekinese
n02086240-Shih-Tzu
n02085620-Chihuahua
True label
48 3 25 11 17
7 65 26 6 23
13 10 120 11 19
10 4 33 47 16
6 7 60 17 62
120
100
80
60
40
20
n02085782-Japanese_spaniel
n02085936-Maltese_dog
n02086079-Pekinese
n02086240-Shih-Tzu
n02085620-Chihuahua
Predicted label


Confusion Matrix of KNN on test data
n02085782-Japanese_spaniel
n02085936-Maltese_dog
n02086079-Pekinese
n02086240-Shih-Tzu
n02085620-Chihuahua
15 4 17 3 9
9 16 16 8 9
7 4 45 7 16
6 2 13 10 8
7 4 31 6 14
40
30
20
10
True label
Predicted label


Multiclass ROC Curve Logistic Regression (Test Data)
Class 0 (AUC = 0.67)
Class 1 (AUC = 0.56)
Class 2 (AUC = 0.56)
Class 3 (AUC = 0.68)
Class 4 (AUC = 0.56)
Micro-average (AUC = 0.63)
Multiclass ROC Curve Logistic Regression (Train Data)
Class 0 (AUC = 0.70)
Class 1 (AUC = 0.73)
Class 2 (AUC = 0.62)
Class 3 (AUC = 0.71)
Class 4 (AUC = 0.61)
Micro-average (AUC = 0.68)
True Positive Rate
False Positive Rate


Multiclass ROC Curve KNN (Test Data)
Class 0 (AUC = 0.63)
Class 1 (AUC = 0.59)
Class 2 (AUC = 0.62)
Class 3 (AUC = 0.62)
Class 4 (AUC = 0.57)
Micro-average (AUC = 0.63)
Multiclass ROC Curve KNN (Train Data)
Class 0 (AUC = 0.91)
Class 1 (AUC = 0.90)
Class 2 (AUC = 0.79)
Class 3 (AUC = 0.85)
Class 4 (AUC = 0.77)
Micro-average (AUC = 0.84)
True Positive Rate
False Positive Rate