MODEL EVALUATION REPORT

By Blessing Ilesanmi

7th December, 2024

Introduction

The evaluation of machine learning models is essential to understand their performance and guide further improvements. This report presents the evaluation of a classification model, based on its performance on a test dataset. The evaluation includes metrics such as accuracy, precision, recall, F1-score, and confusion matrix. These metrics give a comprehensive understanding of how well the model classifies data across multiple classes.

The evaluation was conducted using the following code and results:

1. **Model Accuracy**: The model was evaluated on a test set of images, where the test accuracy was computed.
2. **Classification Report**: This report provides precision, recall, and F1-score for each class, which are useful for evaluating model performance in multi-class classification tasks.
3. **Confusion Matrix**: The confusion matrix was plotted to visualize the performance of the model in terms of true positive, false positive, true negative, and false negative predictions.

Results

The model's performance is summarized in the following key metrics:

Test Accuracy

- **Test Accuracy**: 33.82% This accuracy indicates that the model correctly classified approximately 33.82% of the test set.

Classification Report

The classification report provides a detailed breakdown of precision, recall, and F1-score for each class. Here are the key findings:

1. **Precision** measures the accuracy of positive predictions. For instance, the precision for class 0 is 0.52, meaning that 52% of the instances predicted as class 0 were actually correct.
2. **Recall** measures the ability of the model to identify all relevant instances. For class 0, the recall is 0.66, indicating that 66% of the actual class 0 instances were correctly identified.
3. **F1-Score** is the harmonic mean of precision and recall. For class 0, the F1-score is 0.58, providing a balance between precision and recall.

Below is the classification report for all classes:

	precision	recall	f1-score	support
0	0.52	0.66	0.58	100
1	0.40	0.42	0.41	100
2	0.19	0.26	0.22	100
3	0.16	0.11	0.13	100
4	0.12	0.06	0.08	100
5	0.24	0.21	0.23	100
6	0.31	0.53	0.39	100
7	0.49	0.28	0.36	100
8	0.39	0.50	0.44	100
9	0.56	0.43	0.49	100
10	0.25	0.21	0.23	100
11	0.32	0.09	0.14	100
12	0.47	0.25	0.33	100
13	0.22	0.34	0.26	100
14	0.30	0.25	0.27	100
15	0.27	0.22	0.24	100
16	0.51	0.31	0.39	100
17	0.45	0.54	0.49	100
18	0.28	0.39	0.33	100
19	0.29	0.24	0.26	100
20	0.61	0.68	0.64	100

21	0.38	0.45	0.41	100
22	0.35	0.24	0.28	100
23	0.42	0.57	0.48	100
24	0.50	0.62	0.56	100
25	0.20	0.32	0.25	100
26	0.26	0.20	0.23	100
27	0.23	0.25	0.24	100
28	0.71	0.46	0.56	100
29	0.78	0.18	0.29	100
30	0.29	0.22	0.25	100
31	0.25	0.30	0.28	100
32	0.33	0.29	0.31	100
33	0.36	0.35	0.36	100
34	0.22	0.24	0.23	100
35	0.23	0.27	0.25	100
36	0.32	0.33	0.32	100
37	0.20	0.36	0.26	100
38	0.13	0.11	0.12	100
39	0.36	0.30	0.33	100
40	0.33	0.28	0.30	100
41	0.52	0.65	0.58	100
42	0.24	0.30	0.27	100
43	0.30	0.42	0.35	100

44	0.30	0.10	0.15	100
45	0.24	0.17	0.20	100
46	0.28	0.15	0.19	100
47	0.38	0.56	0.46	100
48	0.59	0.58	0.59	100
49	0.38	0.46	0.41	100
50	0.22	0.08	0.12	100
51	0.23	0.30	0.26	100
52	0.62	0.25	0.36	100
53	0.49	0.69	0.57	100
54	0.40	0.34	0.37	100
55	0.11	0.10	0.11	100
56	0.42	0.43	0.43	100
57	0.35	0.22	0.27	100
58	0.40	0.43	0.41	100
59	0.23	0.45	0.31	100
60	0.58	0.81	0.68	100
61	0.50	0.35	0.41	100
62	0.39	0.37	0.38	100
63	0.20	0.29	0.24	100
64	0.12	0.07	0.09	100
65	0.11	0.09	0.10	100
66	0.26	0.27	0.27	100

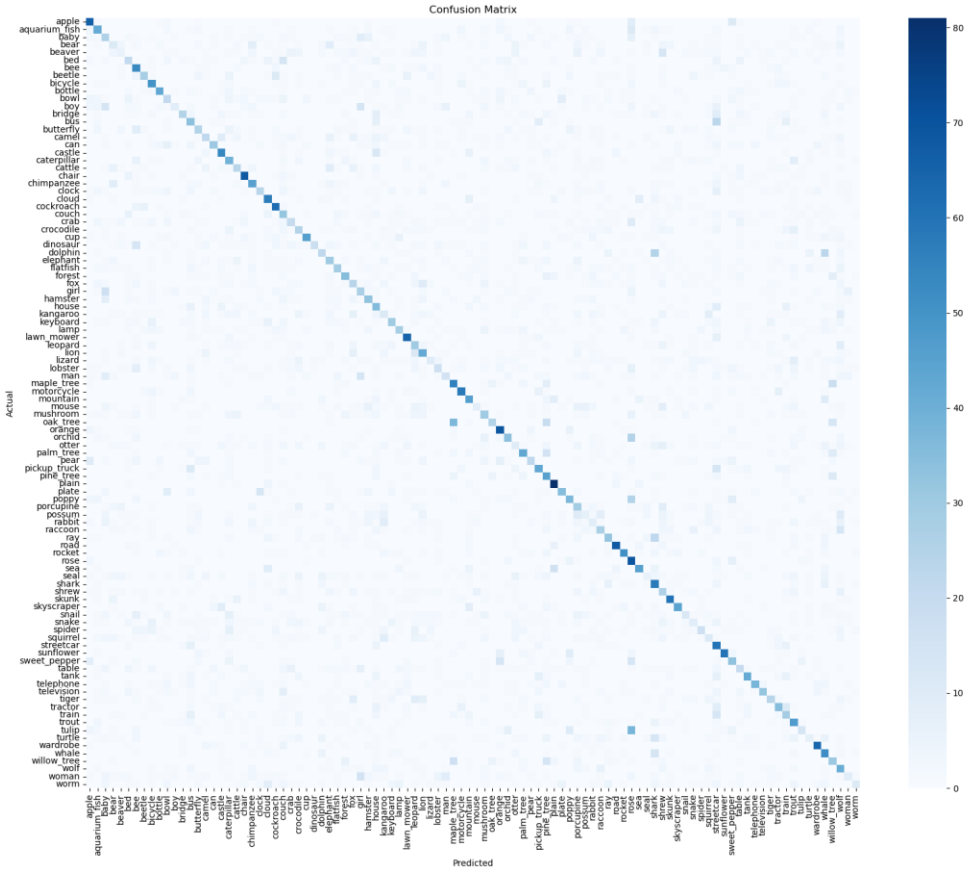
67	0.39	0.32	0.35	100
68	0.67	0.68	0.67	100
69	0.62	0.51	0.56	100
70	0.24	0.69	0.36	100
71	0.62	0.45	0.52	100
72	0.16	0.06	0.09	100
73	0.27	0.58	0.37	100
74	0.17	0.26	0.20	100
75	0.51	0.60	0.55	100
76	0.63	0.44	0.52	100
77	0.27	0.12	0.17	100
78	0.15	0.09	0.11	100
79	0.47	0.17	0.25	100
80	0.13	0.11	0.12	100
81	0.22	0.60	0.32	100
82	0.80	0.60	0.69	100
83	0.29	0.33	0.31	100
84	0.21	0.20	0.21	100
85	0.49	0.42	0.45	100
86	0.64	0.38	0.48	100
87	0.64	0.32	0.43	100
88	0.32	0.24	0.27	100
89	0.37	0.35	0.36	100

90	0.25	0.29	0.27	100
91	0.30	0.47	0.37	100
92	0.33	0.14	0.20	100
93	0.16	0.07	0.10	100
94	0.63	0.64	0.64	100
95	0.29	0.52	0.38	100
96	0.26	0.31	0.28	100
97	0.21	0.41	0.27	100
98	0.24	0.09	0.13	100
99	0.42	0.11	0.17	100
accuracy			0.34	10000
macro avg	0.35	0.34	0.33	10000
weighted avg	0.35	0.34	0.33	10000

The **macro average** and **weighted average** provide an overall evaluation of the model's performance across all classes. The weighted average considers the support (number of instances per class) in its computation, while the macro average treats all classes equally.

Confusion Matrix

A confusion matrix was generated to visualize the performance of the model. Each row of the matrix represents the actual class, and each column represents the predicted class. The matrix provides insight into how well the model is performing and where it is making errors.



Conclusion

Based on the evaluation results, the model achieved a test accuracy of 33.82%, which indicates a significant room for improvement. The classification report shows that the model performs inconsistently across different classes, with some classes having higher precision and recall (e.g., class 0 with precision of 0.52 and recall of 0.66) while others have poor scores (e.g., class 4 with precision of 0.12 and recall of 0.06).