

Name & ID:

Md. Mehdi Hassan Fahim - 1921539042

Ittihad Ahmed Anindya - 2031372642

Kazi Minhaz Uddin Sami - 2031733642

Group No - 01

CSE445.06

Submitted To:

M. Shifat-E-Rabbi

"Create a custom image classification dataset featuring five classes: dog, cow, cat, lamb, and zebra, each with 100 images sourced from the internet or captured using your phone. Develop a classification model to classify these classes with at least 90% accuracy."

Abstract

This report details a project aimed at creating a custom image classification dataset consisting of five classes: dog, cow, cat, lamb, and zebra. Each class contains 100 images sourced either from the internet or captured using a smartphone. A classification model was developed using these images, achieving an accuracy of at least 90%. The project utilized a Random Forest Classifier and various Python libraries to preprocess the images, train the model, and evaluate its performance.

Introduction

Image classification is a critical area in machine learning that involves categorising images into predefined classes based on their content. This project focuses on building a custom dataset featuring five distinct animal classes, enabling the development of an effective image classification model. The objective is to classify images with high accuracy,

contributing to the understanding of machine learning applications in visual recognition tasks. This project not only enhances skills in data collection and preprocessing but also provides practical experience in implementing machine learning algorithms.

Methodology

The project was carried out through the following steps:

- 1. Dataset Creation:** A dataset was compiled, consisting of 100 images for each of the five classes. Images were collected from the internet and personal devices.
- 2. Image Preprocessing:** Images were resized to a uniform dimension (128x128 pixels) and converted into a numerical format suitable for model training. A helper function was implemented to handle loading

and processing the images efficiently.

3. **Data Splitting:** The dataset was split into training and validation sets using an 80-20 ratio to ensure robust model evaluation.
4. **Model Training:** A Random Forest Classifier was selected for its effectiveness in handling image data. The model was trained on the processed image data.
5. **Model Evaluation:** The model's accuracy was evaluated on the validation set, ensuring it met the project requirement of at least 90% accuracy.
6. **Model Saving:** The trained model was saved for future use and deployment.

Features

The key features of the project include:

- **Custom Dataset:** Five animal classes with diverse images, allowing for a balanced classification task.
- **Image Preprocessing:** Resizing and flattening images for effective model training.
- **Random Forest Classifier:** Utilizing a robust machine learning algorithm suitable for classification tasks.
- **Model Evaluation:** Validation accuracy metrics to ensure performance standards are met.
- **Model Persistence:** Saving the trained model for future predictions and applications.

Libraries Used

The following libraries were used to implement the project:

- `os`: For file handling and directory operations.
- `numpy`: For numerical operations and handling image data as arrays.
- `sklearn.model_selection`: For splitting the dataset into training and validation sets.
- `sklearn.ensemble`: For implementing the Random Forest Classifier.
- `sklearn.metrics`: For evaluating model accuracy.
- `tensorflow.keras.preprocessing.image`: For image loading and preprocessing.
- `collections.Counter`: For counting the distribution of classes in the dataset.
- `joblib`: For saving the trained model to disk.

Summary

In conclusion, the project successfully created a custom image classification dataset and developed a classification model that achieved an accuracy of over 90%. By utilizing effective methodologies for data collection, preprocessing, and model training, this project demonstrates the potential of machine learning in image recognition tasks. The use of a Random Forest Classifier highlights the flexibility and robustness of ensemble learning methods in achieving high-performance metrics in classification problems. The trained model is saved and ready for future applications, marking a significant accomplishment in the field of machine learning and computer vision.