Fruit Classification Project: A Machine Learning Approach

Master's Degree in Artificial Intelligence
Department of Computer Science
University of Verona
Academic Year 2024

Author: Hailegabrel Dereje Degefa

Fruit Classification Project: A Machine Learning Approach

Hailegabrel Dereje Degefa

Academic Year 2024

1 Introduction

This project aims to classify different types of fruits using machine learning techniques. The primary objective is to develop a model capable of accurately categorizing fruit images into their respective classes. The project employs various feature extraction techniques and classification algorithms to achieve this goal.

2 Motivation

The motivation behind this project is to enhance the accuracy of fruit classification systems, which can be beneficial in agricultural automation, retail, and quality control processes. Accurate fruit classification can improve sorting and packaging processes, reduce human error, and enhance the overall efficiency of fruit handling operations.

3 Dataset

The dataset comprises images of various fruit types, collected from a diverse range of sources to ensure variability and robustness in the classification model. The dataset is divided into training and testing sets, with 80% used for training and 20% for testing.

3.1 Dataset Preparation

1. **Data Collection**: Images of different fruits are collected and organized into respective directories.
2. **Data Augmentation**: Techniques such as rotation, flipping, and zooming are applied to increase the dataset's variability.
3. **Loading and Preprocessing**:
    - **Resizing**: Images are resized to a uniform dimension of 64x64 pixels.
    - **Normalization**: Pixel values are scaled to a range of 0 to 1.

4 Methodology

4.1 Feature Extraction

- **Principal Component Analysis (PCA)**: Reduces the number of features by retaining the most significant components.

4.2 Model Selection

Various machine learning algorithms are tested to identify the most effective model for fruit classification. These include:

- **Support Vector Machines (SVM)**

- **K-nearest neighbors (KNN)**
- **Naive Bayes**

5 Evaluation and Results

The performance of each classifier is evaluated using accuracy, precision, and F1 score. The evaluation is conducted in two phases: before applying PCA and after applying PCA.

5.1 Experimental Results

Table 1: Experimental Results

Model	**Accuracy (%)**	**Precision**	**F1 Score**
SVM (only Normalization)	92.1	92.3	92.0
KNN (only Normalization)	85.5	86.0	85.4
Naive Bayes	88.7	89.1	88.6
SVM (PCA)	94.6	94.8	94.5
KNN (PCA)	89.4	89.7	89.3
Naive Bayes (PCA)	90.2	90.5	90.1

6 Conclusions

- The experimental results demonstrate the effectiveness of various machine learning models in fruit classification.
- Applying PCA improves the classification accuracy of the models.
- Support Vector Machines (SVM) consistently achieved the highest accuracy across different scenarios, followed by Naive Bayes and K-nearest neighbors (KNN).
- These findings indicate that SVM could be the preferred choice for fruit classification tasks due to its robust performance.

7 Recommendations

- **Expand the dataset**: Collect more samples of various fruit types to improve the robustness and generalization of the model.
- **Explore advanced techniques**: Implement Convolutional Neural Networks (CNN) and other deep learning architectures for potentially better performance.
- **Experiment with more classifiers**: Besides SVM, Naive Bayes, and KNN, explore other algorithms such as Gradient Boosting Machines or Neural Networks to find the most effective approach for fruit classification.