Plant Species Classification: High Accuracy with Traditional ML

This project successfully developed a robust plant species classification model achieving 92.06% accuracy, adhering to a strict constraint: no transfer learning or pre-trained deep learning models were used. This demonstrates the enduring power of feature engineering and classical machine learning in complex classification tasks.

Project Constraints & Key Achievement

Primary constraint was to develop a high-accuracy classification model without leveraging transfer learning or pre-trained Convolutional Neural Networks (CNNs) like ResNet.

Despite this, I have achieved a remarkable 92.06% accuracy using only traditional machine learning methods.

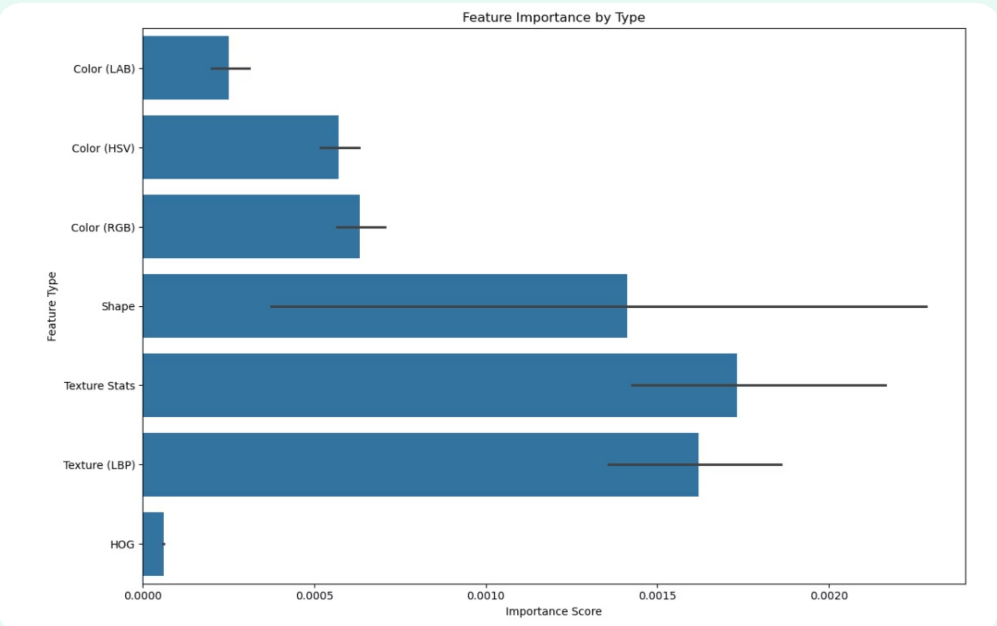
This accomplishment highlights the efficacy of meticulously engineered features coupled with classical ML algorithms, proving their competitive viability even in an era dominated by deep learning.

Methodology: Handcrafted Features & Classical ML

In this my approach prioritized handcrafted feature extraction, avoiding neural networks entirely , extracted four distinct types of features:

- **Color:** RGB, HSV, and LAB histograms, totaling 1,536 dimensions, capturing intricate color distributions.
- **Texture:** Local Binary Patterns (LBP) combined with statistical measures (mean, standard deviation), essential for surface pattern recognition.
- **Shape:** Contour-based features like area and circularity, providing geometric descriptors.
- **HOG:** Histograms of Oriented Gradients, capturing structural and edge patterns, contributing 3,780 dimensions.

The classical ML pipeline involved initial data preprocessing with **SimpleImputer** for missing values and **StandardScaler** for feature scaling. The core model was a **Random Forest classifier** (200 trees, max depth 15), chosen for its robustness and ability to handle high-dimensional feature spaces effectively.



Problem Statement

Manual plant species identification is a time-consuming process that demands specialized botanical expertise. My goal was to automate this classification from images, strictly limiting ourselves to feature engineering and traditional machine learning. I utilized a carefully curated dataset with distinct training and testing splits.

Results & Insights

The model achieved a test accuracy of **92.06%**, demonstrating that traditional ML, when applied rigorously, can indeed achieve performance comparable to simpler deep learning approaches without transfer learning. The confusion matrix revealed minimal misclassifications, indicating high reliability across various species.

Feature importance analysis highlighted color and HOG features as the strongest predictors, which aligns intuitively with how botanists often classify plants based on leaf color, shape, and venation patterns.

Challenges & Solutions

- **Challenge:** Avoiding transfer learning posed a risk of lower accuracy.
- **Solution:** I mitigated this with aggressive, multi-faceted feature engineering across four distinct types, combined with the power of ensemble learning (Random Forest).
- **Challenge:** The handcrafted features resulted in a high-dimensional feature space.
- **Solution:** Random Forest's inherent feature selection capabilities and scalability were crucial in managing this complexity efficiently.

Broader Implications & Future Work

This project stands as a significant achievement, proving that traditional ML can yield competitive results. Its implications extend to:

- **Resource-Constrained Environments:** Validates traditional ML for applications on edge devices or in low-power computing scenarios where deep learning models might be too computationally expensive.
- **Interpretability:** Offers clear interpretability through feature importance, contrasting with the "black-box" nature of many deep learning models.

Future work includes exploring hybrid approaches combining handcrafted features with shallow neural networks and optimizing the model for mobile deployment using lightweight ML libraries like TensorFlow Lite.