



Automated Leaf Classification using Machine Learning and Image Processing

Md Nahin Islam
U00899990
Date: 23rd April, 2025



Introduction

- Importance in botany, agriculture, environmental monitoring
- Manual identification is time-consuming and requires expertise
- Aim: Use AI to automate species recognition from leaf images



Objectives

- Build a dataset of leaf species
- Apply preprocessing & feature extraction
- Train SVM, Random Forest, and CNN models
- Evaluate models using standard metrics
- Develop a real-time identification system



Dataset Overview

- UCI Leaf Dataset
- Contains images of 40 plant species
- Two types of data:
 - RGB images for CNN-based modeling
 - CSV file with 14 pre-extracted features (shape, texture, etc.) for SVM & RF



Methodology

- SVM
 - Used 14 numerical features from CSV
- Random Forest
 - Ensemble of decision trees
- CNN
 - Trained directly on RGB images
 - Included data augmentation to reduce overfitting

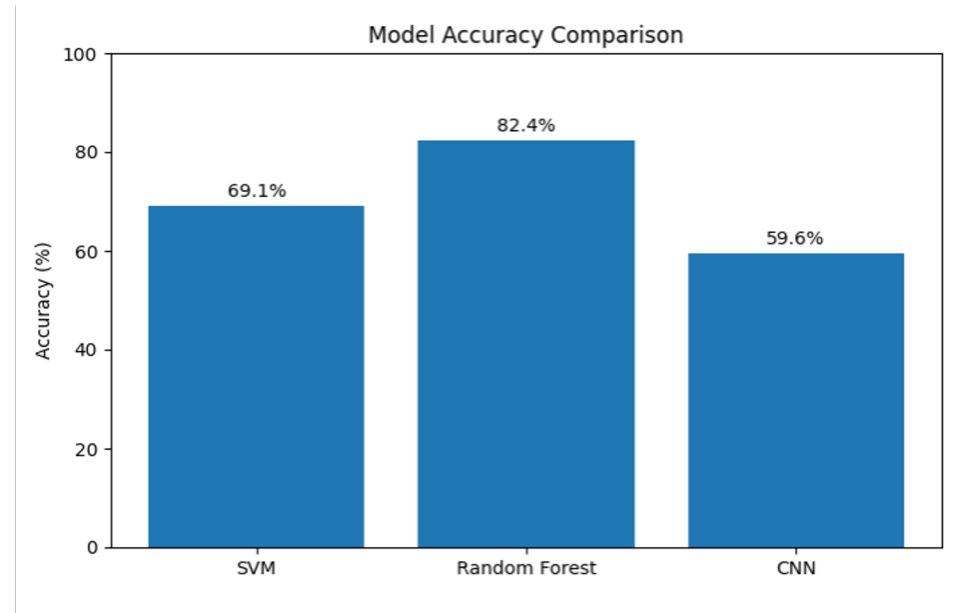


Data Preprocessing

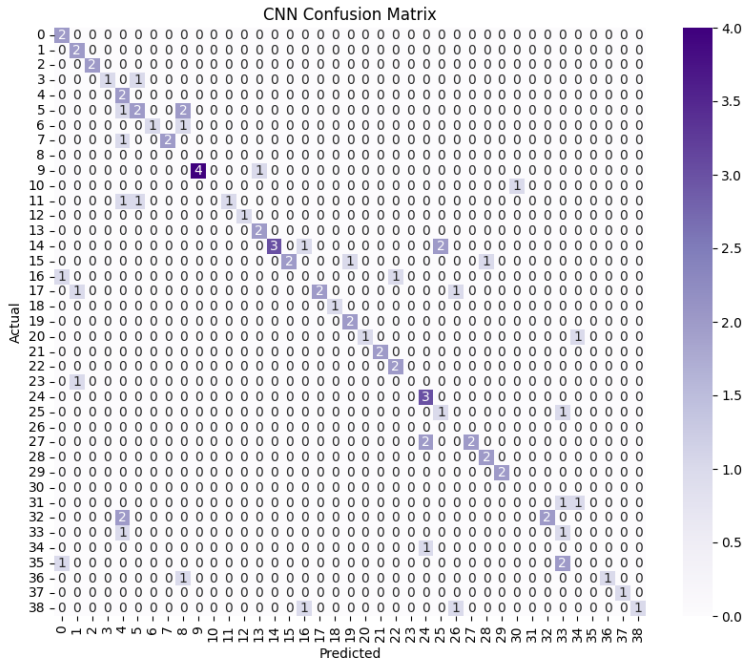
- For SVM & Random Forest (CSV Data)
 - Applied StandardScaler to normalize feature values
- For CNN
 - Resize all images to 128x128 pixels
 - Applied Data Augmentation to combat overfitting

Results & Model Comparison

- Random Forest: 82.4%
(Best Performer)
- SVM: 69.1%
- CNN: 71%



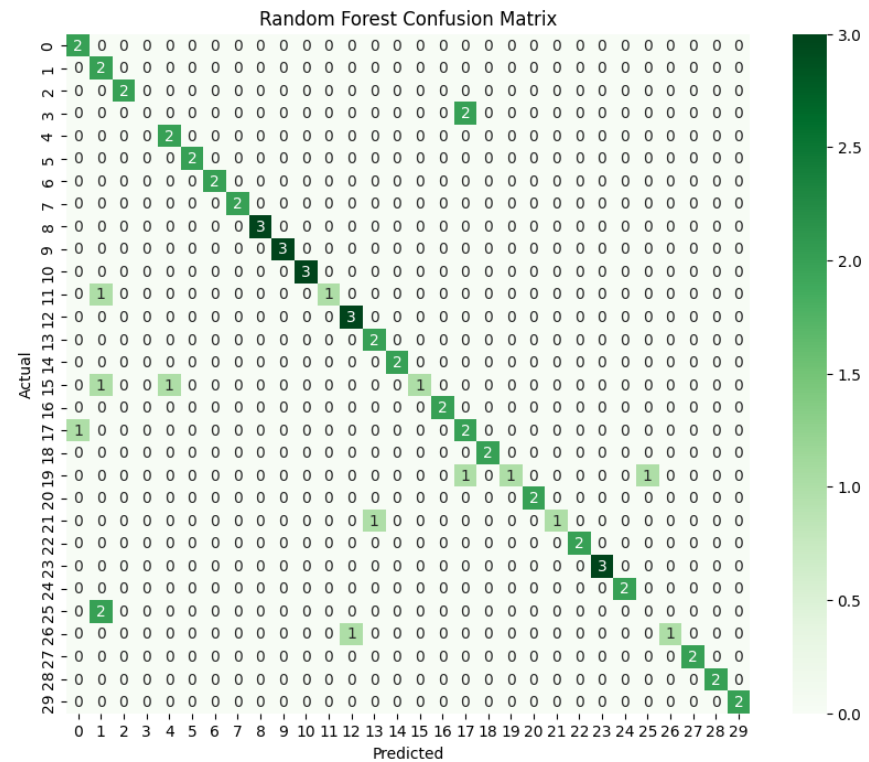
-



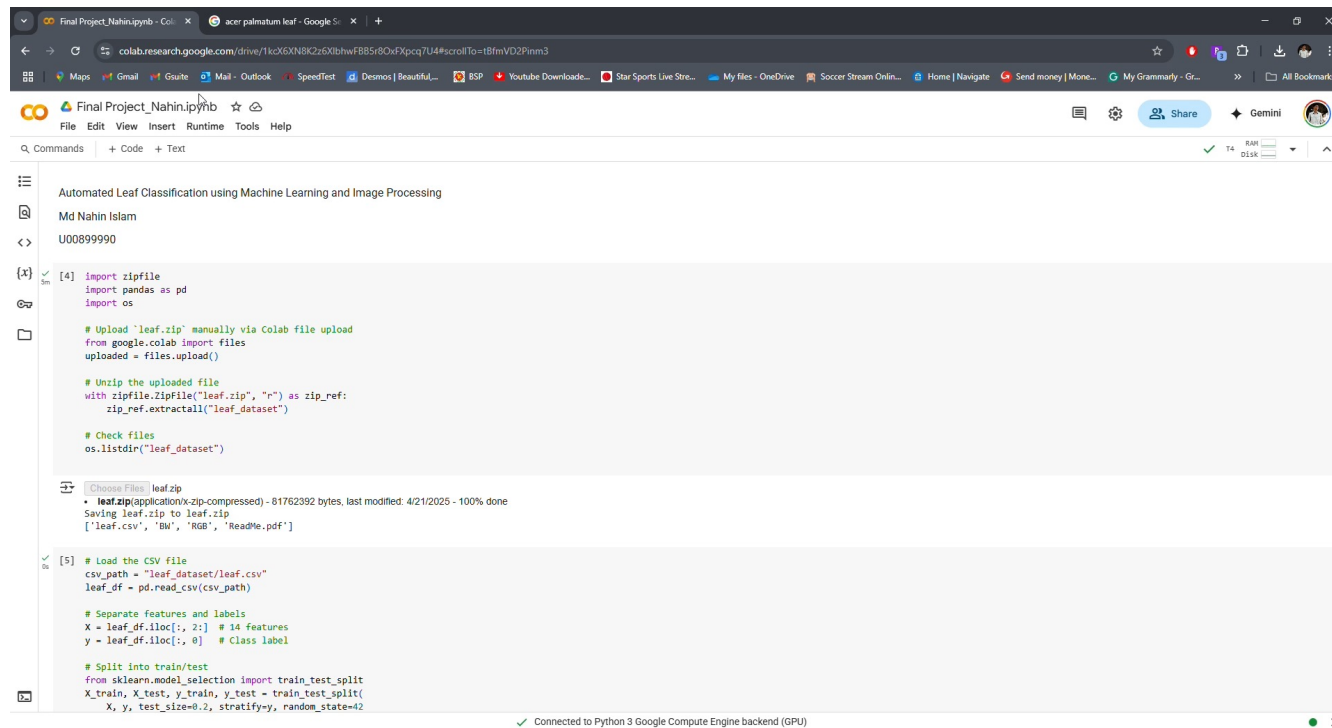


Results & Model Comparison

- Random Forest
Confusion Matrix



Real-Time Leaf Identification System



The screenshot displays a Google Colab notebook interface. The browser address bar shows the Colab URL. The notebook title is "Final Project_Nahin.ipynb". The code is organized into two cells. Cell [4] contains code for uploading a file, unzipping it, and checking its contents. Cell [5] contains code for loading a CSV file, separating features and labels, and splitting the data into training and testing sets. The status bar at the bottom indicates "Connected to Python 3 Google Compute Engine backend (GPU)".

```
[4] import zipfile
import pandas as pd
import os

# Upload 'leaf.zip' manually via Colab file upload
from google.colab import files
uploaded = files.upload()

# Unzip the uploaded file
with zipfile.ZipFile("leaf.zip", "r") as zip_ref:
    zip_ref.extractall("leaf_dataset")

# Check files
os.listdir("leaf_dataset")

[5] # Load the CSV file
csv_path = "leaf_dataset/leaf.csv"
leaf_df = pd.read_csv(csv_path)

# Separate features and labels
X = leaf_df.iloc[:, 2:] # 14 features
y = leaf_df.iloc[:, 0] # Class label

# Split into train/test
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, stratify=y, random_state=42
```



Conclusion & Future Work

- Implemented SVM, Random Forest, and CNN for automated leaf classification
- Random Forest achieved highest accuracy using structured features
- CNN performance improved to 71% with data augmentation techniques
- Developed a real-time identification system
- Expand dataset with diverse, real-world leaf images

THANK
YOU!
