PlantDoc: A Dataset for Visual Plant Disease Detection

Davinder Singh, Naman Jain, Pranjali Jain, Pratik Kayal, Sudhakar Kumawat, Nipun Batra
Indian Institute of Technology Gandhinagar

Introduction

India loses 35% of the annual crop yield due to plant diseases.

Early detection of plant diseases remains difficult for farmers due to the lack of necessary expertise.

Computer vision approaches can enable scalable, cost-effective and early plant disease detection.

PlantVillage Dataset

54,306 images
14 crop species and 26 diseases
Diseased(D) images of 11 species
Healthy(H) images of 12 species

Images collected under **controlled laboratory** conditions
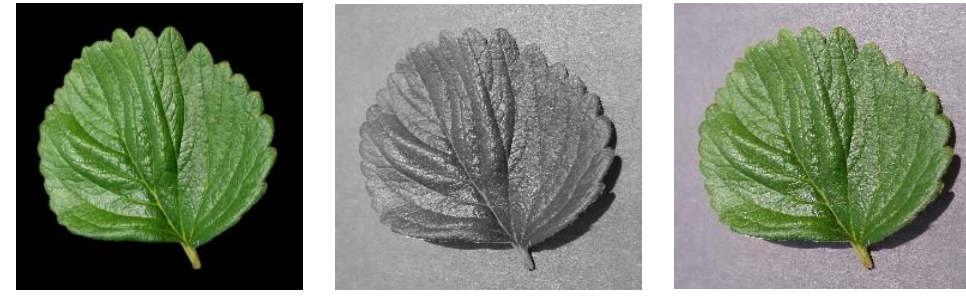
Research Gap

Images in PlantVillage dataset (PVD) are taken in **laboratory setups**.

PVD limits the effectiveness of detecting diseases using image processing techniques in the real cultivation fields.

Plant images may contain **multiple leaves** with different types of **background conditions** and **varying lighting conditions.**

PlantDoc Dataset

2,598 images
Color Images from Real-Life Farms
13 crop species (17 D and 10 H classes)
Diseased images - 13 species
Healthy images - 10 species
Downloaded from Google and Ecosia
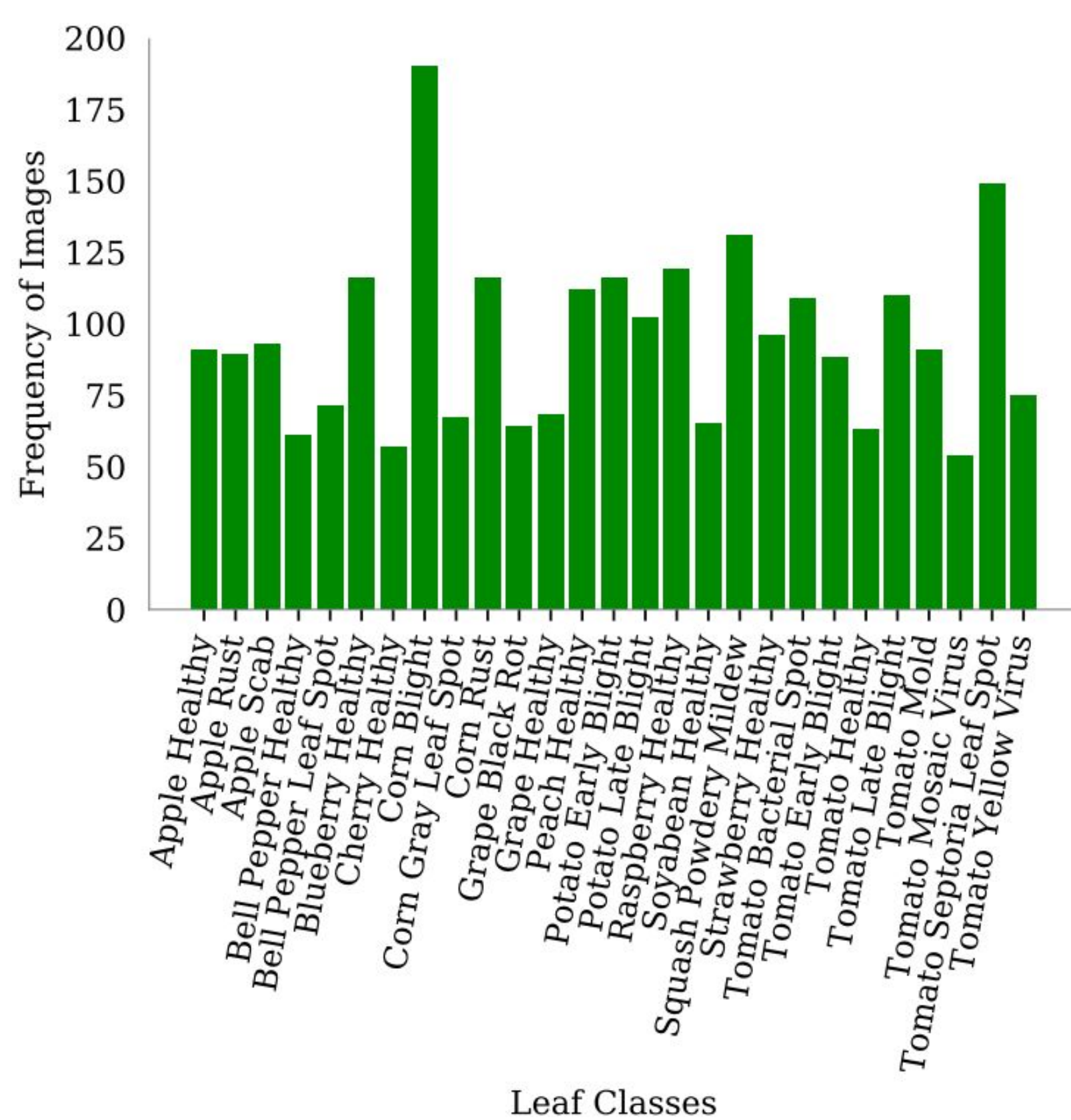
20,900 Total images downloaded → APS Filtering → **2,598** PlantDoc Dataset → Bounding boxes → **9,216** Cropped-PlantDoc

Plant Image Classification

PreTrained Weights	Training Set (Set %)	Test Set (Set %)	Accuracy	F1-Score
ImageNet	PlantDoc (80)	PlantDoc (20)	13.74	0.12
ImageNet+PVD	PlantDoc (80)	PlantDoc (20)	**29.73**	**0.28**

PVD helps the model in learning about leaves

| Model	PreTrained Weights	Training Set (Set %)	Test Set (Set %)	Accuracy	F1-Score
VGG16	ImageNet	C-PD (80)	C-PD(20)	44.52	0.44
VGG16	ImageNet	PVD	C-PD (100)	19.73	0.18
VGG16	ImageNet+PVD	C-PD (80)	C-PD (20)	60.41	0.60
InceptionV3	ImageNet	C-PD (80)	C-PD (20)	46.67	0.46
InceptionV3	ImageNet	PVD	C-PD (100)	30.78	0.28
InceptionV3	ImageNet+PVD	C-PD (80)	C-PD (20)	62.06	0.61
InceptionResNet V2	ImageNet	C-PD (80)	C-PD (20)	49.04	0.49
InceptionResNet V2	ImageNet	PVD	C-PD (100)	39.87	0.38
InceptionResNet V2	ImageNet+PVD	C-PD (80)	C-PD (20)	**70.53**	**0.70**

PVD cannot be directly used in detecting diseases on real-life images

Leaf Detection

Model	PreTrained Weights	mAP (at 50% iou)
MobileNet	COCO	32.8
MobileNet	COCO+PVD	22.4
Faster-rcnn-inception-resnet	iNaturalist	36.1
Faster-rcnn-inception-resnet	COCO	**38.9**

PVD is not contributing towards better results in object detection architecture

Results and Discussion
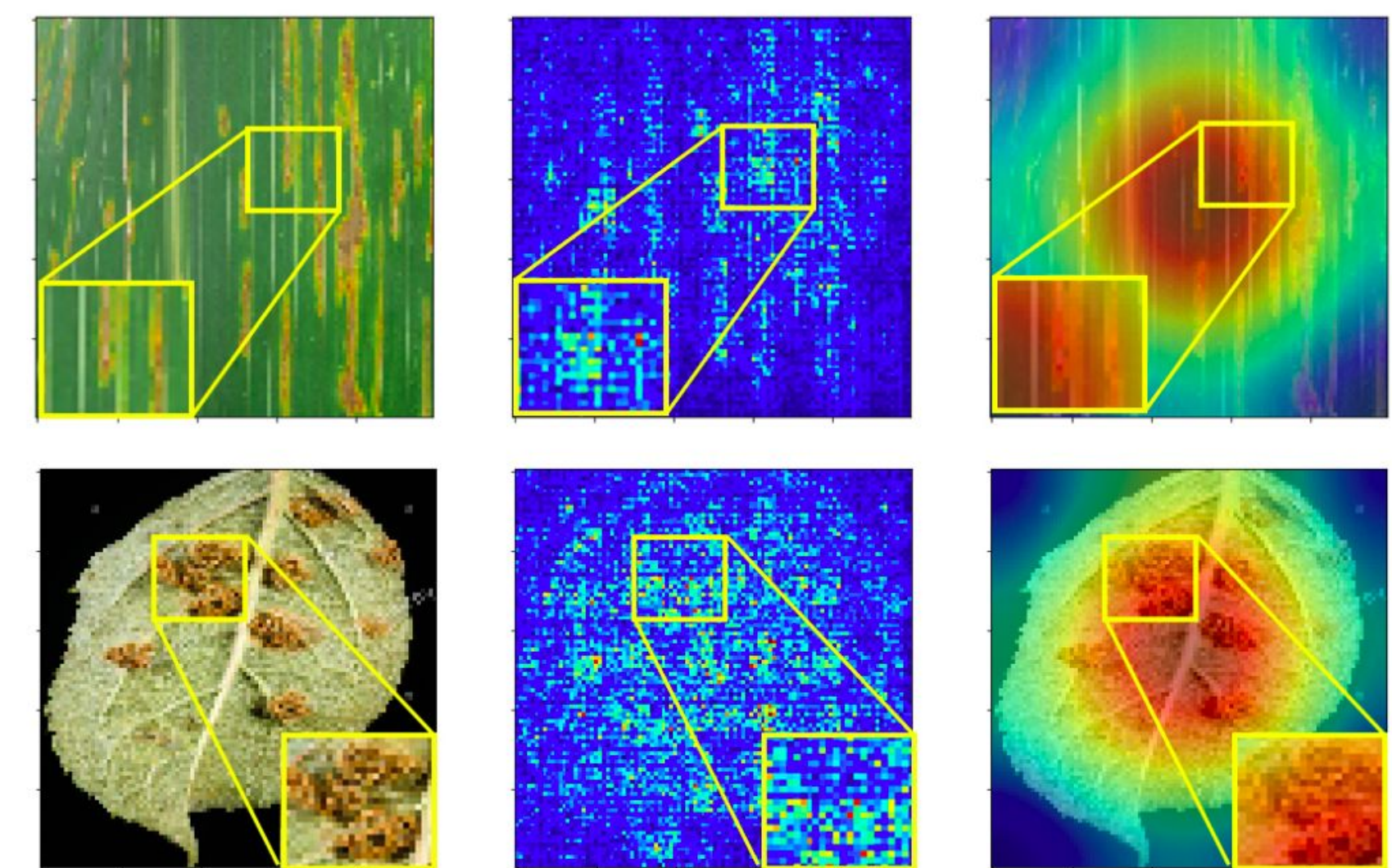
The model focuses on the **shape of the leaf** to distinguish between different species and the **set of visual features correlated with the plant disease** such as the blemishes on the leaf.

Datasets like PlantDoc enable scalable plant disease detection through computer vision which incorporates leaf images in natural environmental conditions.

Limitations

Lack of extensive domain expertise

Small size of the dataset

Dataset