*Semi-supervised Leaf Pathology Classification*

Singh, Manmeet  
Charles W. Davidson College of Engineering  
San Jose State UniversitySan Jose, California  
manmeet.singh01@sjsu.edu

Vadgama, Chinmay  
Charles W. Davidson College of Engineering  
San Jose State UniversitySan Jose, California  
[chinmay.vadgama@sjsu.edu](mailto:chinmay.vadgama@sjsu.edu)

line 1: 3rd Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCID  
  
line 1: 4th Given Name Surname  
line 2: *dept. name of organization (of Affiliation)*  
line 3: *name of organization (of Affiliation)*line 4: City, Country  
line 5: email address or ORCI

***Abstract*—This e**

***Keywords—component, formatting, style, styling, insert (****key words****)***

# Introduction

Misdiagnosis of the many diseases impacting agricultural crops can lead to misuse of chemicals leading to the emergence of resistant pathogen strains, increased input costs, and more outbreaks with significant economic loss and environmental impacts. Current disease diagnosis based on human scouting is time-consuming and expensive, and although computer-vision based models have the promise to increase efficiency, the great variance in symptoms due to age of infected tissues, genetic variations, and light conditions within trees decreases the accuracy of detection.

We explore plant disease detection by following computer vision based deep learning architectures on a large plant image dataset. We explore traditional supervised architectures such as ResNet-50, InceptionResNetV2 etc. and compare it with new self-supervised learning approaches such as SimCLR architectures. Generally there is very little labeled dataset available for plants and specific diseases. Using self-supervised techniques such as SimCLR could provide ground for using large unlabeled dataset on top of learning from small labeled dataset. New data can be collected with help of drone based cameras without any manual help and that data could be used as an unlabeled source for SimCLR network. This could get us more generalization and model adaptation can be increased even if we don’t have higher accuracy than supervised techniques.

# Related Work

# Data

## OASIS-1

* 1. Alzheimer’s severity
  2. Plant Village
  3. Plant Pathology

# Methods

For the semi-supervised learning approach, we used simclrv2 which builds on top of contrastive learning based simclr. The contrastive approaches use heavy data augmentation followed by learning of the non linear transformation between representation and the contrastive loss which substantially improves the quality of learned representations. Lastly, the contrastive learning benefits from larger batch sizes compared to supervised learning. By combining these methodologies, we are able to achieve high performance with small amounts of labeled data.

Specific to simclrv2, a large amount of unlabeled data is used for unsupervised pre-training followed by supervised fine-tuning. This paradigm uses unlabeled data in a task-agnostic way initially. The key ingredient of this approach are the deep and wide networks during pretraining and fine-tuning. The fewer the labels in this approach, the more the pre-training benefits from this approach.

In our first approach, we used simclrv2 to perform unsupervised pre training on the OASIS-1 dataset. Next, we fine-tuned using the Alzheimer's severity dataset from Kaggle. After exhausting the possibilities for this approach, we pivoted to pre-training with plant village dataset and fine-tuned on the plant pathology dataset.

We trained our models on a combination of using the P100 GPU on the HPC machine, as well as the Google Colab environment.

# Experiments

# Conclusion

##### References