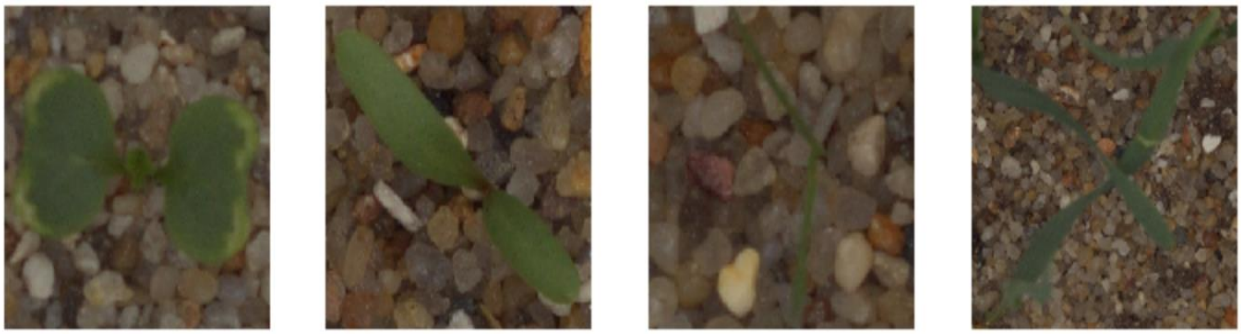
PLANT SEEDLING CLASSIFICATION

Domain Background

- Agriculture is vital for human survival and remains a major driver of several economies around the world; more so in underdeveloped and developing economies. Plants continue to serve as a source of food and oxygen for all life on earth.
- In continents like Africa, where agriculture is predominant, proper automation of the farming process would help optimize crop yield and ensure continuous productivity and sustainability.
- The transformation of the agricultural sector by use of smart farming methods can power economic growth in many countries. According to [1], there is a strong link between increased productivity and economic prosperity.

One major reason for reduction in crop yield is weed invasion on farmlands. Weeds generally have no useful value in terms of food, nutrition or medicine yet they have accelerated growth and parasitically compete with actual crops for nutrients and space. Inefficient processes such as hand weeding has led to significant losses and increasing costs due to manual labour [2].



Precision agriculture, with the goal of defining systems that support decision-making in farm management in order to optimize returns on outputs while preserving resources, and weed control systems have been developed aiming at optimizing yields and costs while minimizing

environmental challenges; some robotic systems have been used to do this [3]. The robots and the vision machines need to be able to precisely and reliably detect a weed from the useful plants. Machine vision technologies developed for selective weeding face a challenge of reliable and accurate weed detection. It's not easy to identify the weeds due to unclear crops boundaries, with varying rocky or sandy backgrounds, and as a result, traditional classification methods are likely to fail on this task [4].

In this work, I am trying to explore the performance of traditional computer vision methods on this task and show that a Deep Convolutional Neural Network (CNN) does the best job at classifying plant seedlings. In computer vision, CNNs have been known to be powerful visual models that yield hierarchies of features enabling accurate segmentation. They are also known to perform predictions relatively faster than other algorithms while maintaining competitive performance at the same time [5].

[1] http://www.fao.org/3/I9900EN/i9900en.pdf

[2] Y. Gharde, P. Singh, R. Dubey, and P. Gupta, *Assessment of yield and economic losses in agriculture due to weeds in India*, Crop Protection, vol. 107, pp. 12–18, 2018.

[3] http://pure.au.dk/portal/files/114969776/

[4] https://www.cs.swarthmore.edu/

[5] W. Shang, S. Kihyuk, A. Diogo, and L. Honglak, *Understanding and improving convolutional neural networks via concatenated rectified linear units*, International Conference on Machine Learning, pp. 2217-2225. 2016.

INSPIRATION:

My inspiration is due to the reason,With increasing demand for food and cash crops, due to a growing global population and the challenges posed by climate change, there is a pressing need to increase farm outputs while incurring minimal costs.

PROBLEM STATEMENT:

The aim of the project is to find the CNN-driven seedling classification applications, when used in farming automation has the potential to optimize crop yield and improve productivity and efficiency when designed.

DATASETS AND INPUT:

Explore plant seedling classification dataset from Kaggle at the below link provided

Dataset Link: https://www.kaggle.com/c/plant-seedlings-classification

This dataset contains 4750 images of plant species.

Dataset containing images of approximately 960 unique plants belonging to 12 species at several growth stages.

Each image has a filename that is its unique id. The goal is to create a classifier capable of determining a plant's species from a photo. Test set we need to predict the species of each image.

We performed image preprocessing on the dataset before training the models. we smoothen the image, removing high frequency content and then converting this blurred version to HSV space. We created a mask by specifying a range of possible color values of the seedlings to be captured and using a morphological erosion with an 11x11 structuring kernel, we are able to produce foreground seedling images with the backgrounds subtracted. We then perform image normalization by subtracting the mean from each pixel, dividing by the standard deviation and then scaled the data in the range of [0,1].

SOLUTION:

CNNs have been widely used for diverse image classification tasks. My solution is to attempt to use keras and CNN for this problem and finally I am going to find the accuracy of the predicted model.

BENCHMARK MODEL:

In this plant seedling classification, I want to set worst benchmark model by adding only one convolution Layer and Max pooling layer and dense layers with 2 units.

Then I will try to improve the benchmark model accuracy by adding sufficient layers to the model.

EVALUATION METRICS:

I want to use Accuracy as a evaluation metric which will give the (correct predictions/total number of predictions) and lost metrics,since our model consists of 12 plant species.

PROJECT DESIGN:

STEP 1:-**Loading the train,validation,test datasets**

- Since while downloading the dataset ,it is already splitted into train and test set.
- I will store these train,test datasets into some directory variables.

STEP 2:-Create a Sequential Model

- For the benchmark model ,I implement with non neural networks techniques and come up with CNN model and then we read the training data images and resize the all the images into 128*128.Then I create a model with activation ReLU and in last layer add a “softmax”.The softmax function id=s often used in the final ayer of a neral network-based classifiers.

STEP 3:-Visualization Results

- After the completion of the model, ,i will visualize the training set results as well as testing set results.