**Project Report**

**Project Title**

**Plant Seedling Classification using Neural Nets**

**Team Leader:**

A Harshith 21R11A0504

**Team Members:**

P Rohith 21R11A0537

G Harshit 21R11A0519

G Pranav 21R11A0515

D Nissi 21R11A0511

**INTRODUCTION**

There were two image datasets are provided which are used for training and testing while developing the model. The datasets contain set of images of plant seedlings at various stages of grown. Each image has a filename that is its unique id. The dataset comprises 12 plant species.

The species are as follows:

* Black-grass
* Charlock
* Cleavers
* Common Chickweed
* Common wheat
* Fat Hen
* Loose Silky-bent
* Maize
* Scentless Mayweed
* Shepherds Purse
* Small-flowered Cranesbill
* Sugar beet

1. Black-grass: A type of grassy weed commonly found in agricultural fields. It competes with crops for nutrients and can reduce crop yields if not managed properly.
2. Charlock: Another common weed in agricultural areas, charlock is a member of the mustard family. It has yellow flowers and can be a nuisance for farmers as it competes with crops for resources.
3. Cleavers: Also known as sticky weed or goosegrass cleavers are fast-growing weeds with tiny hooked hairs that cling to surfaces. They can smother crops if left unchecked.
4. Common Chickweed: This is a widespread weed with small, star-shaped flowers. It thrives in moist soil and can quickly spread in gardens and fields.
5. Common wheat: A staple cereal grain widely cultivated for its edible seeds. Wheat is a major crop used for making flour, bread, pasta, and other food products.
6. Fat Hen: A broadleaf weed that competes with crops for nutrients and water. It has dense clusters of green flowers and can be problematic in fields and gardens.
7. Loose Silky-bent: A grassy weed with fine, silky hairs on its leaves. It can invade grasslands and reduce the quality of pasture for grazing animals.
8. Maize: Also known as corn, maize is a cereal grain domesticated by indigenous peoples in the Americas. It is widely cultivated for food, animal feed, and industrial products.
9. Scentless Mayweed: A common weed with daisy-like flowers and a strong, unpleasant odour when crushed. It can be found in a variety of habitats, including fields, roadsides, and waste areas.
10. Shepherd's Purse: A winter annual weed with heart-shaped seed pods resembling purses. It can be a problem in agricultural fields and disturbed areas.
11. Small-flowered Cranesbill: Also known as "geranium," this weed has small, delicate flowers and deeply lobed leaves. It can quickly spread and compete with crops for resources.
12. Sugar beet: A root crop cultivated for sugar production. Sugar beet is processed to extract sugar and is an important source of sucrose for the food industry.

**Methodology**

The methodology included following stages while developing the model: Importing the packages

Importing the datasets.

Pre-processing the datasets.

Building the image classifier using CNN. Evaluating the image classifier.

Training Dataset: This dataset contains images of various plant species, organized into subfolders corresponding to each species. The training dataset is utilized for developing the predictive model.

Testing Dataset: This dataset consists of images of plants and is used to evaluate the performance of the trained model by predicting the plant species.

To ensure effective training and evaluation of the model, a portion of the images from the training dataset was allocated for validation purposes. The dataset splitting was performed using a ratio of 80% for training and 20% for validation.

Split Ratio:

Training Dataset: 80%

Test Dataset: 20%

## Pre-processing the datasets

1. Data Augmentation: Data augmentation technique is applied to create variations of the original images. The following augmentation techniques are applied:
   * rotation\_range: Random rotation of the image by a specified angle range here we did 20 degrees.
   * width\_shift\_range and height\_shift\_range: Randomly shifting the width and height of the image by a fraction of the total width and height.
   * shear\_range: Applying shear transformations to the image.
   * zoom\_range: Randomly zooming into or out of the image.
   * horizontal\_flip: Flipping the image horizontally.
   * brightness\_range: Adjusting the brightness of the image within the specified range.
   * fill\_mode: Strategy used for filling in newly created pixels resulting from transformations.
2. Rescaling: The rescale parameter is set to 1./255, which rescales pixel values from the range [0, 255] to the range [0, 1]. This normalization step ensures that the input values to the neural network are within a similar range, which can help improve convergence during training.
3. Flow from Directory: The flow\_from\_directory method is used to flow images from a directory structure. It automatically infers class labels from subdirectory names.

## Images after loading Training dataset





* + The above figures display the first five images of the loaded training dataset and there are 3800 images.

## Images after loading Validation dataset





* + The above figures display the first five images of the loaded validation dataset and there are 947 images.

## Images after loading Testing dataset

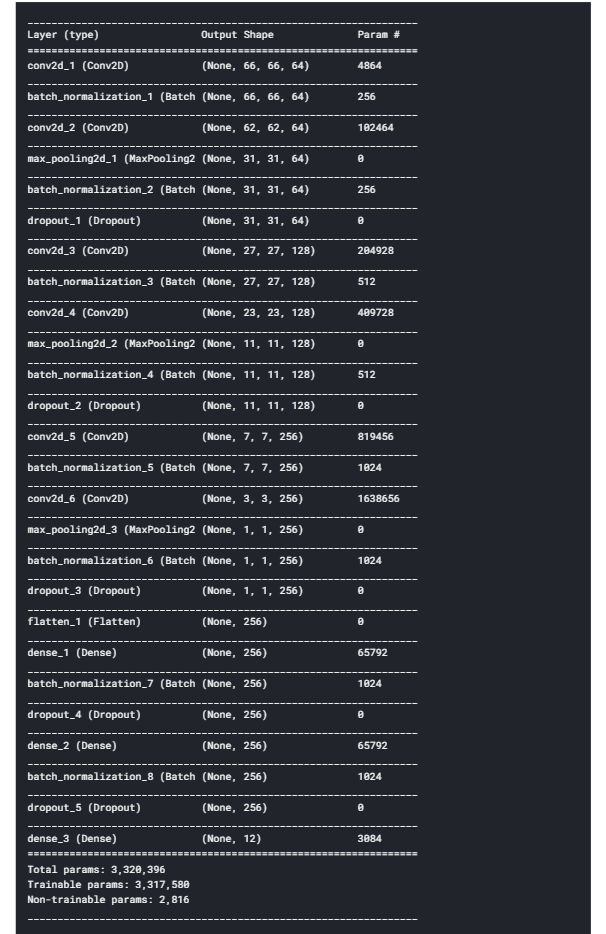




* + The above figures display the first five images of the loaded testing dataset and there are 794 images.

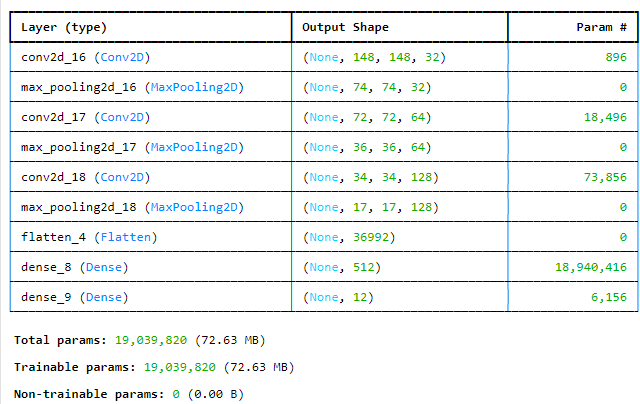
As we have to specify or mention what is the sampling for the test dataset so there is no class is represented.

DEFINING THE CONVOLUTIONAL NEURAL NETWORK

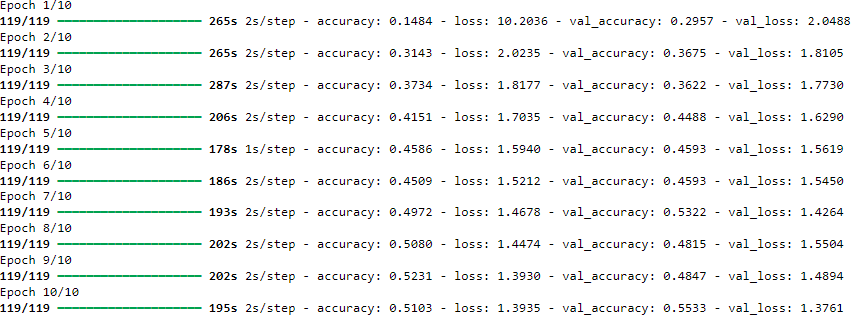


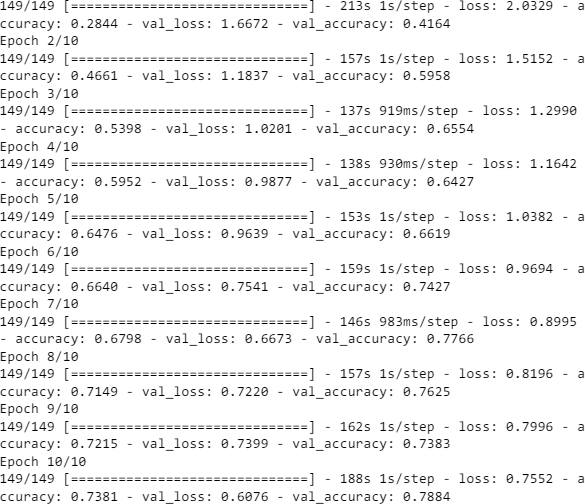
**FITTING THE CNN ONTO THE DATA**

* + Here we have to define the model architecture, including convolutional layers for feature extraction, max-pooling layers for down sampling, and fully connected layers for classification.
  + Finally, the model is compiled using categorical cross entropy loss, the Adam optimizer, and accuracy as the evaluation metric.



* + The model is trained using the fit () method, where training data is provided through training dataset, and validation data is provided through validation dataset. The model is trained for 10 epochs.
  + The history object stores training/validation loss and accuracy values for each epoch, allowing for performance evaluation and visualization.



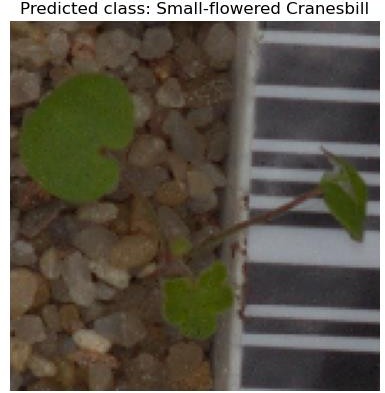
* + The above figure shows the evaluation metrices.
  + As the above accuracy is very low, we have done Data Augmentation where Augmenting training data with transformations such as rotation, shifting, flipping, and zooming can help increase the diversity of your dataset and improve the generalization of our model.
  + The above figure shows the evaluation metrices where it is has accuracy of 0.78.
  + The trained CNN model is used to predict the class labels for the images in the test data.



* + These are predicted class names for the plants in test dataset.
  + As there are 794 images to predict in test dataset. We have given output for each class prediction only.

## OUTPUTS

**Prediction of Small flowered Cranesbill**



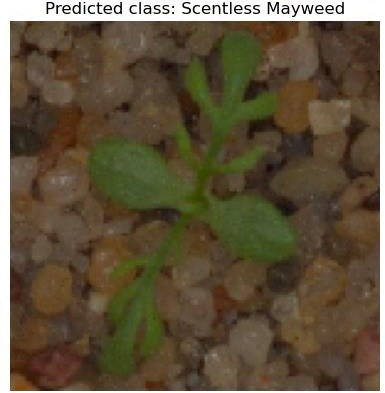
## Prediction of Shepherds Purse



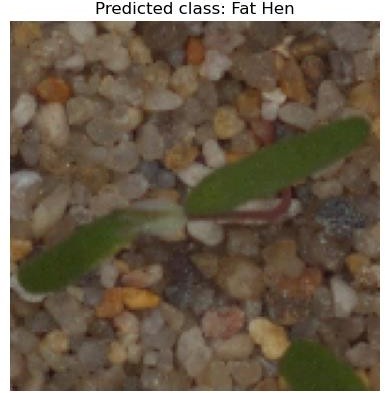
**Prediction of Charlock**



## Prediction of Scentless Mayweed



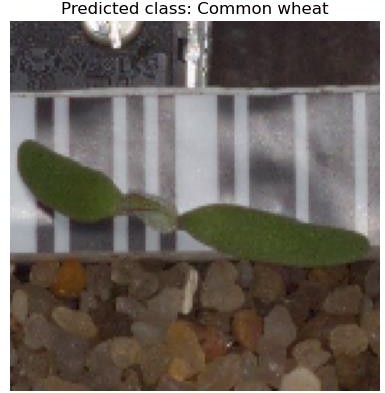
**Prediction of Fat Hen**



## Prediction of Cleavers



**Prediction of Common wheat**



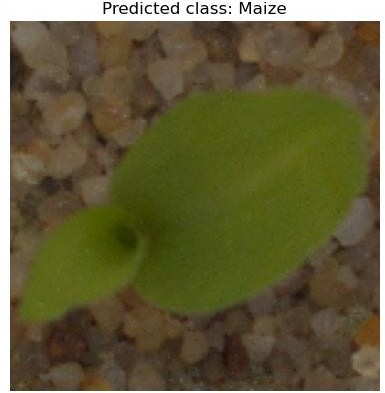
## Prediction of Common Chickweed



**Prediction of Loose Silky-bent**



## Prediction of Maize



**Prediction of Sugar beet**



## Prediction of Black-grass

The Black grass plant sampling is not predicted across 794 images. The model has not predicted for this sampling because it may not be included in test dataset or our model has to be more accurate to predict.

The model has predicted for 749 images in the dataset. We have shown particularly image data class and images for all type of samplings. We are showing the approximately 16 images from 749 images which are in below.

These are the predicted class labels and images.



**CONCLUSION**

In conclusion, the CNN model developed for plant seedling classification has shown promising results in predicting various plant species with an overall accuracy of 0.78 after implementing data augmentation techniques. While the model accurately predicted the majority of plant species in the testing dataset, it failed to predict instances of Black grass due to possible absence in the test dataset or the need for further refinement of the model. Despite this limitation, the model demonstrates potential for efficient and precise classification of plant seedlings, offering a valuable tool for agriculturalists in identifying and managing weed species. Continued refinement and expansion of the dataset, along with fine-tuning of the model parameters, could further enhance its performance and utility in real-world agricultural applications**.**