

**NILKAMAL SCHOOL OF MATHEMATICS, APPLIED STATISTICS AND ANALYTICS**

M. Sc. Statistics and Data Science

Semester – III

Research Treatise – III

**Group Members**

Tamay Desai (A006)

Sachin Dwivedi (A008)

Rohit Fuke (A010)

Dhruv Gilda (A012)

Harsh kandalgaonkar (A020)

**Faculty Mentor**

Dr. Kavita Jain

**ACKNOWLEDGEMENT**

We want to convey our heartfelt gratitude to DR. KAVITA JAIN, our mentor, for her invaluable advice and assistance in completing our project. She was there to assist us every step of the way, and her motivation is what enabled us to accomplish our task effectively. We would also like to thank all of the other supporting personnel who assisted us by providing the knowledge that was essential and vital, without which I would not have been able to perform efficiently on this project.

We would also like to thank the SVKM’s NMIMS (NILKAMAL SCHOOL OF MATHEMATICS APPLIED STATISTICS AND ANALYTICS) for accepting our project in the desired field of expertise. We would also like to thank our friends and parents for their support and encouragement as we worked on this Project.

**CONTENTS**

**Page no.**

1. Problem statement…………………………………………………………………….4

2. Motivation………….……………….………………………………………………...4

3. Abstract…………………………………………………………………………….....5

4. Introduction.……………………………………………………………………….… 5

5. Data Description……………………………………………………………………...6

6. Data Pre-processing…………………………………………………………………..8

7. Model Building………………………………………………………………………11

8. Model Evaluation………………………………………………………...…………..14

9. Model Comparison…………………………………………………...………………15

10. Limitations…………………………………………………………………………...15

11. Conclusion………………………………………………………………………..….16

12. Future scope………………………………………………………………………….16

13. Bibliography…………………………………………………………………………17

**PROBLEM STATEMENT**

The global agriculture industry is pivotal for ensuring food production and security on a large scale. Nevertheless, the menace of plant diseases poses a substantial threat to crop yields, resulting in economic setbacks and food shortages.

For effective control measures and mitigation of these adverse effects, it is imperative to promptly and precisely diagnose plant diseases. Conventional approaches to disease identification often hinge on visual inspections conducted by experts. However, this method is prone to being time-consuming, subjective, and may prove impractical when applied on a large scale.

**MOTIVATION**

Crop diseases are a major threat to food security, but their rapid identification remains difficult in states of the Bharat.

The combination of increasing global smartphone penetration and recent advances in computer vision made possible by deep learning has paved the way for smartphone-assisted disease diagnosis.

The approach of training deep learning models on increasingly large and publicly available image datasets presents a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale.

**ABSTRACT**

Detecting and identifying pests and diseases in plants significantly affects a country's agricultural production. Traditionally, farmers and experts rely on manual observation to spot and diagnose diseases in plants, a process that is both time-consuming and costly, leading to imprecise results. To address this challenge, this paper aims to develop a Disease Recognition Model supported by leaf image classification.

The proposed method utilizes image processing techniques in conjunction with a Convolutional Neural Network (CNN) to enhance the detection of plant diseases. CNNs are a type of artificial neural network specifically designed for processing pixel input, making them well-suited for tasks such as image recognition. This approach aims to streamline and improve the accuracy of plant disease detection, offering a more efficient alternative to the conventional, resource-intensive methods currently in use.

**INTRODUCTION**

Agricultural production, a longstanding method of obtaining food, serves as a crucial income source worldwide. Essential not only for humans but also for animals relying on them for food, oxygen, and other necessities, plants play a vital role in our ecosystem. Governments and experts globally are actively taking initiatives to enhance food production, yielding successful outcomes. Plant diseases can impact various parts of a plant, including the stem, leaf, and branch, with types such as bacterial and fungal diseases varying based on factors like climate.

This environmental interconnectedness becomes evident when a plant is afflicted, affecting all living organisms in the vicinity. Food insecurity arises due to insufficient crop output, exacerbated by climate changes. Early detection of plant diseases is crucial to preventing significant crop losses. Farmers must judiciously use insecticides, as excessive pesticides can harm both crops and farmland. Seeking expert advice is essential to avoid the misuse of chemicals.

Researchers have focused on plants to assist farmers and those involved in agriculture. While visible diseases can be easily detected, innovations like automated disease detection tools, relying on deep learning and neural networks, provide precise and quick results. This study employs a Deep Convolutional Neural Network to identify healthy and infected leaves, aiding in the swift detection of plant diseases. The model, trained with images of both healthy and diseased leaves, accurately determines the condition of a plant based on the input leaf. This technological advancement proves beneficial for farmers engaged in both small and large-scale agricultural cultivation.

**DATA COLLECTION & DESCRIPTION**

Data has been collected from the github [Data set](https://www.kaggle.com/datasets/emmarex/plantdisease). The data set contains the sample of three plants .

1. Potato 2) Tomato 3) Peeper Bell

The different types of Disease and samples are listed below , on the basis of below samples we perform CNN model to detect the disease of plants .

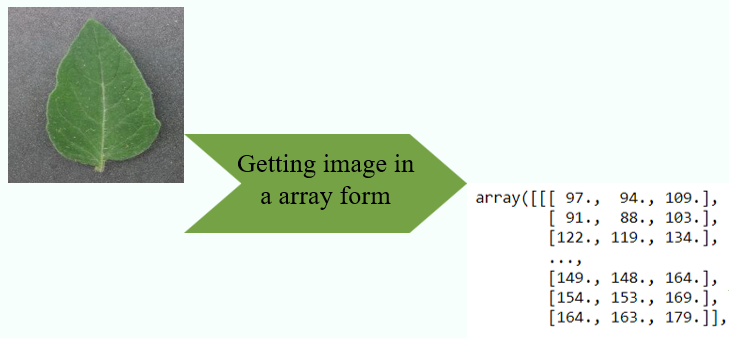
|  |  |  |
| --- | --- | --- |
| Plants | Disease | Sample |
| Potato | Early Blight |  |
|  | Late Blight |  |
|  | Healthy |  |
| Tomato | Early Blight |  |
|  | Late Blight |  |
|  | Target Spot |  |
|  | Mosaic virus, |  |
|  | Yellow leaf curl Virus |  |
|  | Bacterial Spot |  |
|  | Leaf mold |  |
|  | Healthy Tomato |  |
|  | Septoria leaf Spot |  |
|  | Spider Mites |  |
| Peeper bell | Bacterial spot |  |
|  | Bell healthy |  |

**DATA PRE-PROCESSING**

There are certain steps we have taken to process the data before analysis

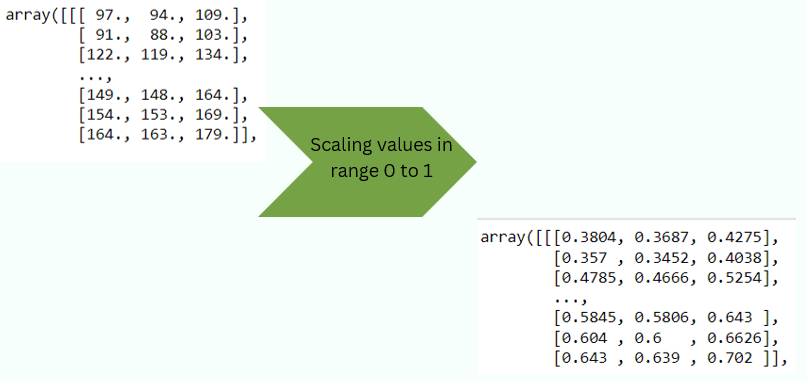
1. Extracting pixel values from images

For extracting pixel values we have defined one function which will return a n-dimensional (here n=256) array which will contain all pixel values and storing those in a list for further analysis.



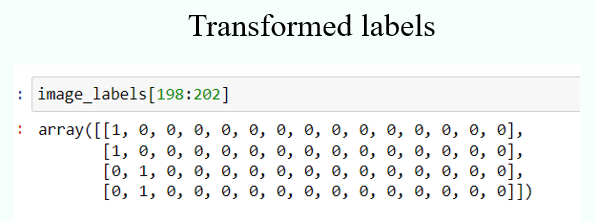
1. Scaling the values within range 0 to 1

Since the pixel values which we have obtained from the previous step contain the pixel values which are ranging between 0 to 255 and we have to scale them for our analysis and we do it by dividing the array by 255 which scales each value to 0 to 1.



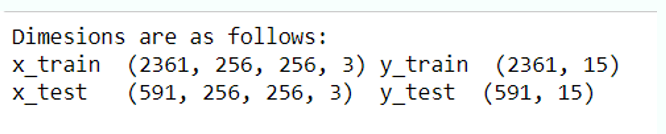
1. Transforming labels using LabelBinarizer

Since we have a data set with labels for each image, we need to transform it into a suitable form for an analysis and we do that using the LabelBinarizer class. It returns a one dimensional array of size 15 (since we have 15 classes).



1. Splitting of dataset into 80-20%

For further analysis, we split our data set into an 80-20 pattern. It means 80% of the data will be used for model training purposes and 20% of the data will be used for validation or testing purposes.

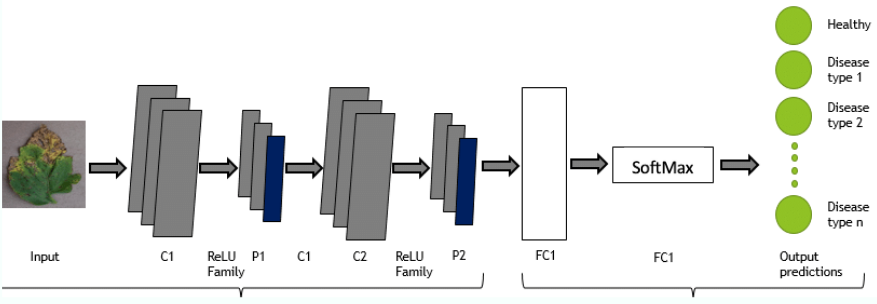


**Convolution Neural Network (CNN)**

Convolutional Neural Networks (CNNs) are a type of artificial neural network designed for processing and analyzing visual data. They are particularly effective in tasks such as image recognition, object detection, and image classification. CNNs have become a fundamental component in computer vision and have demonstrated remarkable success in various applications

* CNNs are a class of deep neural networks designed for processing grid-like data, such as images.
* CNNs have demonstrated exceptional performance in image recognition tasks, surpassing traditional computer vision methods.
* They can automatically learn hierarchical features from images, allowing for robust and accurate classification

Architecture Of CNN:



Here are the key components and concepts of CNNs:

* Convolutional Layers:
  + CNNs use convolutional layers to automatically and adaptively learn spatial hierarchies of features from input images. Convolution involves sliding small filters (kernels) over the input image to detect patterns like edges, textures, and more complex structures.
* Pooling Layers:
  + Pooling layers (commonly max pooling) are used to downsample the spatial dimensions of the feature maps obtained from convolutional layers. This reduces the computational complexity and makes the network more robust to variations in input.
* Activation Functions:
  + Non-linear activation functions, such as ReLU (Rectified Linear Unit), are applied element-wise after convolutional and pooling operations. ReLU introduces non-linearity to the model, enabling it to learn complex patterns.
* Fully Connected Layers:
  + After convolutional and pooling layers, the network often includes one or more fully connected layers. These layers connect every neuron to every neuron in the preceding layer. Fully connected layers are responsible for learning global patterns and relationships.
* Flattening:
  + Before the fully connected layers, the output from the previous layers is flattened into a one-dimensional vector. This is necessary because fully connected layers require one-dimensional input.
* Dropout:
  + Dropout is a regularization technique used in CNNs to prevent overfitting. It randomly drops a fraction of neurons during training, forcing the network to learn more robust features.
* Softmax Activation:
  + In classification tasks, the output layer often uses the softmax activation function to produce probabilities for each class. This allows the network to make predictions and assign the input to one of the predefined classes

**Model Building**

Here we consider different layers of model of CNN

1. 5 Convolutional Layer and 1 Hidden Layer.
2. 5 Convolutional Layer and 2 Hidden Layer.
3. 4 Convolutional Layer and 2 Hidden Layer.

Some of the model parameters

Optimizer: ADAM

Learning Rate: 0.001

Metrics: Accuracy

Batch Size: 32

Epochs: Ranges from 17 - 25

**Model 1: 5 Convolutional Layers and 1 Dense Layer**

line graph of training and validation accuracy for a model with five convolutional layers and one dense layer. The training accuracy is higher than the validation accuracy, which is a sign of overfitting. However, the gap between training and validation accuracy is relatively small, which suggests that the model is not severely overfitting. The training accuracy is around 73% and the validation accuracy is around 70%. This means that the model is able to correctly predict around 73% of the training examples and around 70% of the validation examples. This is a good performance for this type of model, but there is still some room for improvement.

The fact that the model is performing better on the validation set than on the test set suggests that the model is generalizing well to new data.

The model is relatively simple, with only five convolutional layers and one dense layer. This suggests that the model is not overfitting the training data to a large extent.

The model is able to achieve a good accuracy on a challenging task, which suggests that the model is effective at learning the patterns in the data.

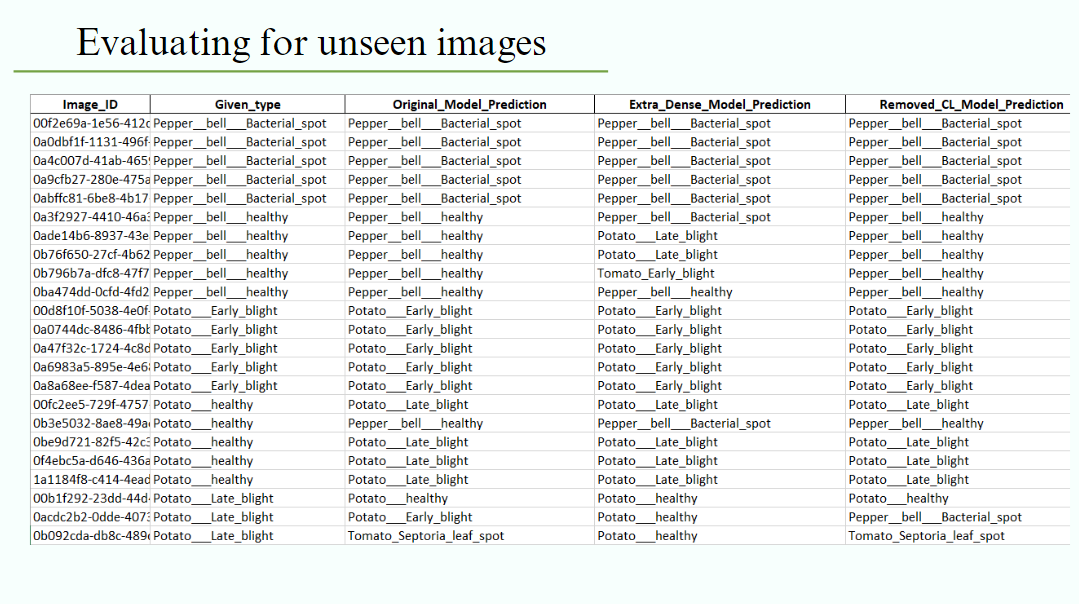
**Model 2: 5 Convolutional Layers and 2 Dense Layer**

Here is this same thing happening for 5 convolution layers and 2 dense layers our test accuracy is 76.31133 % which is pretty much good

**Model 3: 4 Convolutional Layers and 2 Dense Layer**

Here, for model with 4 convolutional layers and 2 dense layers model is giving test accuracy as 66.83%

**Model Evaluation**



We conducted evaluations on a set of images from the original dataset that were not included in the training phase. These images had predefined classes, referred to as 'Type' in the provided output. All three models were applied to these unseen images, and their predictions proved to be highly accurate, aligning well with the specified classes, as illustrated in the output above.

**Model Comparison**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model | Number of Convolution layer | Number of Dense Layer | epoch | Validation Accuracy |
| 1 | 5 | 1 | 17 | 73.44% |
| 2 | 5 | 2 | 21 | 76.31% |
| 3 | 4 | 2 | 19 | 66.83% |

We have performed many different models with different convolution layers and dense layers but only got one best model with validation accuracy of more than 75% , so we decided to go with it.

We have deployed the above second model for demonstration

**LIMITATIONS**

Some limitations of Convolutional Neural Network (CNN) models for plant disease detection include:

* Limited Data Representations: CNNs heavily rely on the training data they are exposed to. If the dataset used for training doesn't cover a wide variety of plant diseases, the model may struggle to accurately identify less common diseases or new manifestations.
* Generalization Issues: CNNs may face challenges in generalizing well to unseen data or variations in environmental conditions. Factors such as lighting, camera angles, and different plant varieties could impact the model's performance.
* Need for Large Datasets: Training deep learning models like CNNs often requires large and diverse datasets. Acquiring and annotating such datasets for every plant disease can be time-consuming and resource-intensive.
* Computational Resources: Training CNNs, especially deeper architectures, demands substantial computational resources. This can be a limitation for organizations or individuals with limited access to high-performance computing.
* Interpretability: CNNs are often considered as black-box models, making it challenging to interpret the decision-making process. Understanding why a particular diagnosis was made can be crucial, especially in applications like agriculture.
* Robustness to Environmental Changes: CNNs might struggle with robustness to environmental changes. For instance, variations in lighting conditions or image quality could affect the model's accuracy.
* Overfitting: Like many machine learning models, CNNs are susceptible to overfitting, where the model performs well on the training data but fails to generalize to new, unseen data.

More limitation that we faced

* The hardware issue: CNN requires best hardware system to train the model and to evaluate it if the hardware is not as per the requirements, then model will take more time to train on the data set
* Machine learning algorithms were not implemented due to insufficient allocating of the space.
* We have restricted our model to low pixel image.

**CONCLUSION**

By using trial and error method for different convolution layers and dense layers we got the model which efficiently classifies and predicts plant disease with 76.31% accuracy, which means that out of every 100 images it will predict 76 images properly and it will misclassify 24 images.

By using this model, we have successfully deployed the model using API.

**FUTURE SCOUPE**

* Model can be extended to test all the plant's leaves for different fruits and vegetables.
* High quality image can be used to train the model
* If Medicine is available then it can be suggested with it.
* Mobile applications can be made.
* This model can be extended to the Government's kisan application .

**BIBLIOGRAPHY**

H. Park, J. S. Eun and S. H. Kim, “Image-based disease diagnosing and predicting of the crops through the deep learning mechanism”, In Information and Communication Technology Convergence (ICTC), IEEE 2017 International Conference on, pp. 129-131, 2017

K. Elangovan and S. Nalini, “Plant disease classification using image segmentation and SVM techniques”, International Journal of Computa- tional Intelligence Research, vol. 13(7), pp. 1821-1828, 2017.  
 S. H. Lee, C. S. Chan, S. J. Mayo and P. Remagnino, “How deep learning extracts and learns leaf features for plant classification”, Pattern Recognition, vol. 71, pp. 1-13, 2017.

<https://github.com/topics/plant-disease-detection>

<https://www.analyticsvidhya.com/blog/2023/02/plant-disease-classification-using-alexnet/>

<https://www.researchgate.net/publication/357621950_PLANT_DISEASE_DETECTION_USING_IMAGE_PROCESSING_AND_MACHINE_LEARNING_ALGORITHM>