Grenze International Journal of Engineering and Technology, June Issue

Machine Learning Approach for Onion Leaf Disease Detection: A Case Study in Maharashtra, India

Ms. Madhavi Amondkar[1] and Dr. Sachin Bhoite[2]
[1]Vidyalankar School of Information Technology, Department of Information Technology & Data Science, Mumbai, India
Email: mak882017@gmail.com
[2]Dr. Vishwanath Karad, MIT World Peace University, Department of Computer Science & Applications, Pune, India
Email: Sachin.bhoite@mitwpu.edu.in

***Abstract*—Maharashtra is a remarkable state in India that is a major producer of onions, accounting for a sizeable portion of the country's total production. An ideal climate for onion growth is provided in areas like Nashik, Ahmednagar, and Pune Despite its importance, growing onions is filled with difficulties, the most significant of which being the frequency of illnesses that damage onion crops. Serious risks to onion production include diseases which are downy mildew, purple blotch, bacterial blight, white rot & basal rot which result in output losses and financial hardship for farmers. Machine learning models have become extremely effective tools for managing and forecasting agricultural crop diseases such as tomato, corn, grape etc. While nearly all crop leaf diseases may be predicted by these algorithms, very little or none has been done for onions in Maharashtra. There is almost no research done focusing on the onion disease found in different parts of Maharashtra. The purpose of the study is to assist farmers in Maharashtra in recognizing various diseases during the cultivation of onions by utilizing a machine learning model that combines CNN with image classification. Using a dataset of plant leaf diseases, this model ran with 95% accuracy. Additionally, given that the suggested model may be trained on any kind of leaf disease picture dataset and assist in disease type prediction, it can be suggested to work on an onion disease dataset as well.**

***Index Terms*— leaf disease, onion disease, basal rot, purple blotch, machine learning, CNN, Maharashtra.**

I. INTRODUCTION

In World, onion is cultivated over an area of 5.7 million ha (Mha) which represents a production of 106.59 million tonne (MT). India ranks first in onion production (26.64 MT) followed by China (24.16 MT) (FAO, 2021) [1]. Onions contributed about 249 billion Indian rupees in the Indian economy in fiscal year 2021. They are also called as crash crops for farmers in Maharashtra. Being the measure Onion growing state in India with districts from western Maharashtra like Nashik, Pune, Ahmednagar & Solapur are the greatest contributors. Onion diseases are caused by various bacteria or fungi. Furthermore, entire onion fields may be destroyed by the diseases' quick spread, endangering both food security and economic stability. Onion harvests are further supported by the state's favourable soil condition and irrigation capabilities, making onions a profitable crop for farmers. The aim of this study is to identify the Onion disease by classifying & analysing images, which will help Farmers in two ways:

1. Finding the type of disease as early as possible.

Grenze ID: 01.GIJET.10.2.564_2

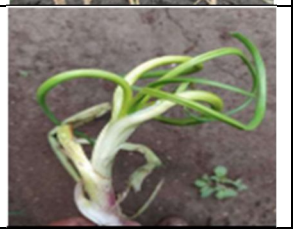
2. Use of pesticides to control the further decaying of the plant.

Now this information will help farmers in growing a good produce with healthy leaves. The research objective is to work on Plant Leaf disease dataset categorise them into Train, Test & Validation, create and run CNN based image classification on the above dataset and check the accuracy of the model, generate confusion matrix. Detailed description about the disease is mentioned below. The outline of the paper is as follows: II. Types of Onion Diseases, III. Literature review, IV. Methodology used, V. Dataset, VI. Results & discussion, VII. Conclusion, VIII. Future scope, IX. References.

II. Onion Diseases

Out of many diseases which are found amongst Onion, the ones which are most prevailing in Western Maharashtra are as follows:

TABLE I: DISEASE CLASSIFICATION (REF: VIKASPEDIA.IN & GOOGLE IMAGES)

Sr. No.	Name of the disease	Scientific Name	Symptoms	Image
1	Purple Blotch	Alternaria porri	On flower leaves & stems that are as tiny, Hollow, white particles with centres that are purple	
2	Basal Rot	Fusarium oxysporum f.sp.cepae	Initially yellowing of leaves & stunted growth of plant is observed & later, the leaves dry from tip downwards	
3	Twister	Collectotrich um blight/anthracnose	Pale yellow, water-soaked patches began to appear on the leaves and then extended lengthwise over the whole leaf blade.	

Diseases mentioned in Table 1, are the most common disease found in Maharashtra. Whereas Purple Blotch is found all over the world. The following table shows the percentage of Onion disease infected districts in Maharashtra [1].

TABLE II: DISTRICT WISE DISEASE PERCENTAGE

Name of the District	% of Disease Incidence
Pune	32.41
Nashik	34.32
Solapur	28.12
Ahmadnagar	30.58

As per their survey shown in Table 2, average disease incidence is 34% with only one disease which is Basal rot. Along with Basal rot, another disease which prevails in the state purple blotch & Twister. Purple blotch is found all over the world, but basal rot is found in an area where Onions are produced continuously. Purple blotch fungus based the diseases whereas basal rot is Pathogen infected disease. Onion plant disease is one of the most key factors due to which the production goes down & quality of the produce is affected. Most of the disease identification or monitoring has been done by leaf observation manually, this method is not only inefficient but also time consuming [11].

Fig 1. Shows county wise onion (Ref: Food and Agriculture Organization of United Nations

III. Literature Review

Many authors from Agricultural background have discussed issues related to production of Onion in Maharashtra & a type of diseases prevailing in the area. Sirisha Thakare et.al in their paper have discussed about how Basal rot is prominent disease found in different Onion growing states in Maharashtra [1]. Another paper which has reviewed Problems in Onion production in Ahmadnagar district in Maharashtra talks about Onion disease as measure challenge. They have suggested if technology can help in some way during the production process so that it may increase the production and can give fair price for it [6]. Another paper written by Immanuelraj et.al have discussed about Maharashtra being the highest growing Onion crop in India, facing issues related to crop yield according to the acreage basis which reduces the yield production [7]. All the above paper suggests some or other intervention of technology either in predicting the type of disease or predicting the crop yield based on analysing numerous factors which are important for plant growth. Several studies that have applied various machine learning and deep learning models to identify plant disease are evaluated to identify the technological aspects of agriculture, particularly in plant disease detection. A lot of research has been done using machine learning algorithms such as CNN, Inception V3, ResNet etc. Few authors have used Plant Village dataset; some have used images from agricultural Universities for testing & predicting a type of disease. Muhammad Ahmad Zali et.al used enhanced IncentionV3 model for classifying a dataset of 1000 images and achieved 85% accuracy in identifying purple blotch as the disease in Onion with the limitation which of using only 1000 images of the Onion plants [3]. Wan-Soo Kim et.al proposed a machine-vision based automatic disease identification which consists of an IOT based model for capturing images and classifying using a pre-trained model [4]. Shima Ramesh et.al have used Random Forest technique in identifying abnormalities that occur in the plant leaves 70% accuracy [8]. A paper by Hiroyuki Takanashi et.al focuses on bacterial infection by rust fungus on Welsh Onions using a model which uses Weibull function. The mathematical model analyses the wetness duration & temperature to identify the severity of infection on Welsh Onions [9]. Another paper by Yo Hong Haw et. Al describes detection of basal rot disease in Plam tree using deep learning model. It has used Fine-tuned DenseNet121as a best performing model with accuracy as 90% [8]. Since the dataset is not readily available, augmentation technique is used by some of the authors such as Jithy Lijo. The focus is to help farmers to enhance the yield by diagnosing disease [10]. Another paper focussed on creating a IoT based mechanism in detecting Onion disease by collecting the real-time information about the soil moisture, temperature, humidity & other related factors affecting the plant health from the sensors which are placed inside a device [11].

IV. Research Methodology

The proposed method makes use of Convolutional Neural Network (CNN) used in machine learning for identifying leaf disease. CNN are specialized neural network architecture used for image classification & object detection or recognition. CNN is made of three important layers. First one being the convolutional layer, next is pooling layer & lastly fully connected layer. Convolutional layer which is the first layer responsible for processing the input whereas the last layer in the model which fully connected layer is responsible for producing the output. Initial layers in the model focusses on identifying the basic information such as colour & edges to identifying different shapes & larger elements present in the image. The journey starts with feature learning till it successfully classifies the image into given category. The convolutional layer makes use of following things -

Input image; b. Filter; c. Feature
Before the images enter the training set it goes through a process which multiplies height, width & depth of the image with weights assigned to it. The pooling layer is in-charge for minimizing the dimensionality of the images with two types of pooling techniques such as max pooling & average pooling shown below.

Max pooling & b. Average Pooling.
Lastly, the fully connected layer is used for performing the classification based on the types of pooling used before that. It further uses ReLU (Rectified Linear Unit) & SoftMax activation function for classification. ReLU is computationally efficient and easy to use in deep learning models especially when it comes to CNN. Softmax function is another effective function used in output layer used in CNN. Both the activation functions are very effective and are mostly used in scenarios where image classification is required.

ReLU & b. Softmax
Architecture of the proposed model is as follows:

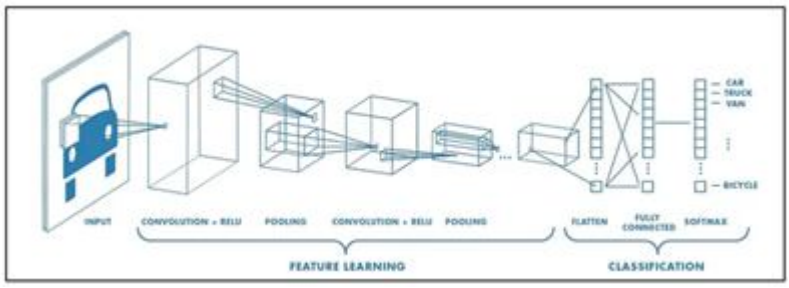


Fig 2. Neural Network with Many Conventional Layers
(Source: Understanding of Conventional Neural Network (CNN) – Deep Learning by Prabhu Raghav-medium.com

Since the dataset for Onion leaf disease was insufficient, we have implemented this model on 'Plant Disease Recognition dataset' which is available for free and has over 1000 leaf images belonging to 3 categories such healthy, rust & powdery mildew. CNN emphasizes fixed size images for all the classes, so either of the size 224x224 or 128x128.

V. DATASET

The proposed model uses Keras libraries for Image classification. The dataset used in this model is Plant Disease Recognition Dataset which is available for free. The dataset has 3 subfolders in it. The abstract view of the folder structure looks like this:

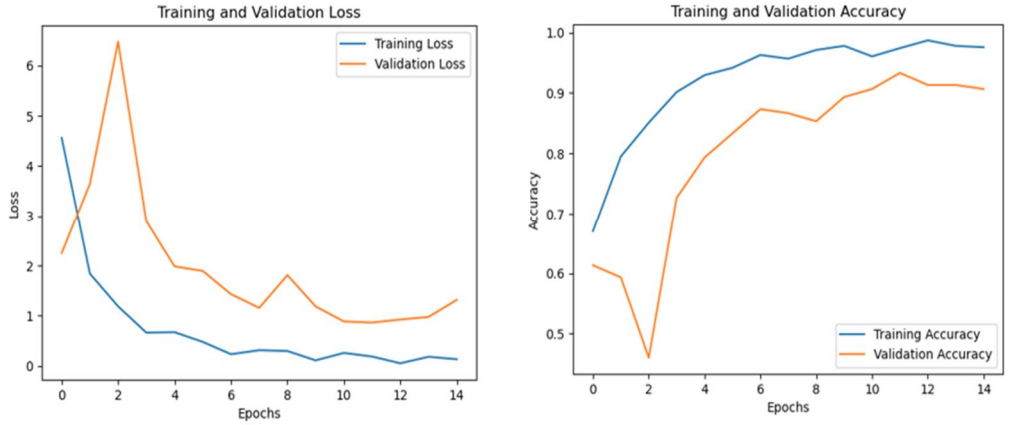
TABLE III: SHOWS STRUCTURE OF THE DATASET USED WITH 3 CLASSES NAMELY: TRAIN, TEST, VALIDATE

Train (1322 Images)	Test (150 Images)	Validation (60 Images)
Healthy	Healthy	Healthy
Powdery	Powdery	Powdery
Rust	Rust	Rust

The above model is trained by dividing the images into 3 classes mentioned above. Rust class has images which are infected with Puccini Ales fungi and caused brown rustic spots on the leaf. The healthy class has all green leaf images. Whereas the Powdery mildew caused by Erysh ales fungi and have powdery layer on the leaves. The training set has 1322 images, Test has 150 images & Validation has 60 images. Overall, this dataset has good quality images for each disease.

VI. RESULTS & DISCUSSION

Training Accuracy & Training Loss: After using the above dataset and implementing CNN from scratch we found Accuracy to be 95%. The accuracy can further be increased by increasing number of epochs. We have taken 15 epochs for this model. Further, we have calculated the confusion matrix, precision, F1 score & recall on the test data. We also have plotted the graph for Training & Validation loss along with Training & Validation accuracy.



Graph: 2 (a) | Graph: 2 (b)

Graph 2(a): Shows the Training & Validation Loss and Graph 2(b) shows the Training & Validation Accuracy. In both the graphs the X-Axis shows the number of Epochs whereas the Y-axis shows the Loss & Accuracy respectively. The increasing Training & Validation shows that the model is training from the data effectively and it is likely to generate good results to unseen data. The decreasing Training & Validation loss also shows that the model is training well and trying extract various features accurately.

Precision & Recall: The value of Precision and recall is generally high if the accuracy is good and even the predictions are also showing correctly. When we tried to find out the precision & recall value of the above model, we found the following results.

```
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 267ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 270ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 265ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 264ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 263ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 265ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 266ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 265ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 268ms/step
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 135ms/step
Train Accuracy  : 97.58 %
Test Accuracy   : 93.33 %
Precision Score : 93.33 %
Recall Score    : 93.33 %
```

Image shows Accuracy, Precision & Recall score

With 93.33% as Precision & Recall score gives us confidence about the model built from scratch. Furthermore, it also implies that it may predict maximum unseen images correctly. It is the proportion of correctly anticipated positive values to all correctly predicted positive values.

a. Precision and recall: These are used to decide the F1-Score. It is considered as the precision and recall harmonic mean.

b. Confusion Matrix: A matrix like representation called a confusion matrix is used to assess the machine learning classification model's performance. The confusion matrix has the following fields in namely True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) are the four types.

c. The true positive predictions are represented by the true positive value.

d. Wrong positive predictions are represented by a False Positive value.

e. The True Negative number denotes accurate negative projection.

f. Inaccurate negative projections are represented as a false negative value.

g. In the confusion matrix, the original class is represented by columns, while the correctly predicted class is represented by rows.

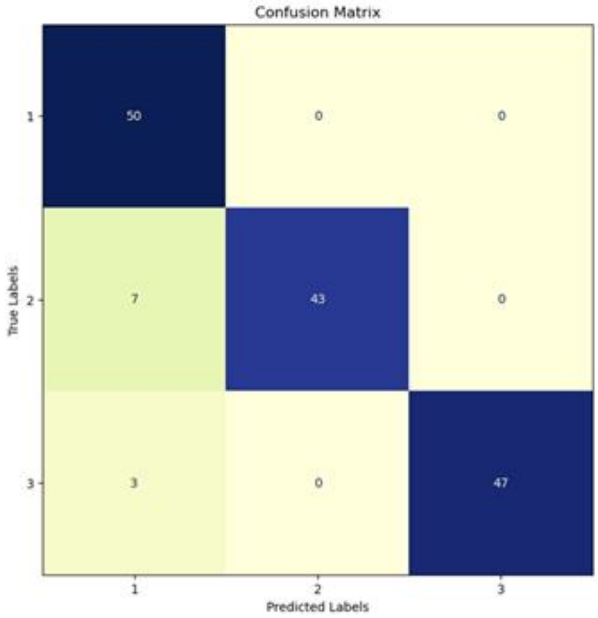


Fig 3. Shows Confusion Matrix

The above matrix shows the complete picture of the model about the number of images that are identified correctly. Predicted labels (1,2,3) stand for Healthy, Rust & Powdery Images. The above Matrix with following order [1,1] identified 50 images correctly, [2,2] identified 43 images correctly & [3,3] has identified 47 images correctly. It shows that though the accuracy is near 100% it is predicting quite a good number of images as per the classes defined.

VII. Conclusion

Technology in agriculture has always shown a remarkable presence and thereby helped farmers and policy makers to take further decisions. Machine learning is one such area which can help Farmers in identifying crop disease. Not only with that but can also help farmers in crop yield prediction for their farms.

The state's onion crop affects Maharashtra's agricultural economy and people's way of life. The persistent occurrence of onion infections poses a danger to the sustainability of onion production. By prioritizing research and innovation in disease prediction and control, stakeholders can ensure food security, safeguard onion harvests, and sustain the socioeconomic well-being of Maharashtra's agricultural communities. Also, identification of diseases at an early stage can help farmers to make use of proper pesticides so this will stop further decaying of the crop and use of adequate pesticides will give less harm to our nature.

Future Scope

Farmers in Maharashtra have been losing income because of infertile soil from improper management of agricultural yield production, irregular rainfall, degraded soil, and low-quality seed. As researchers, we can assist them by providing a technological solution that will enable them to harvest well and profit handsomely. Artificial Intelligence (AI), machine learning, and the Internet of Things (IoT) are three emerging technologies. We can assist farmers in spotting plant diseases by using machine learning. Obtaining the original photos from farms is an issue. We can minimize loss by using machine learning to predict plant diseases at the proper time if we can obtain a small number of high-quality photos of onion leaves that correspond to the disease.

References

[1] Thakare, Sirisha & Dawale, Mukund. (2023). Prevalence and distribution of basal rot disease of onion in major onion growing districts of Maharashtra during Kharif-2022 Sirisha Thakare, SV Kolase, Mukund Dawale, BM Ilhe and AV Chandanshive. The Pharma Innovation. 12. 2812-2815.

[2] Food and Agricultural Organization of United Nations (www.fao.org)

[3] Zaki, M. A., Narejo, S., Ahsan, M., Zai, S., Anjum, M. R., & Din, N. u. (2021). Image-based Onion Disease (Purple Blotch) Detection using Deep Convolutional Neural Network. International Journal of Advanced Computer Science and Applications, 12(5), 448-457.
[4] Wan-Soo Kim, Dae-Hyun Lee, Yong-Joo Kim, Machine vision-based automatic disease symptom detection of onion downy mildew, Computers and Electronics in Agriculture, Volume 168, 2020, 105099, ISSN 0168-1699.
[5] N.K. Agale and B.B. Thavare, Department of Geography, Shri Dnyaneshwar Mahavidyalay Newasa, Chetna Mahavidyalay, Aurangabad Email: agale.sir@gmail.com.
[6] Immanuelraj TK, Dastagiri MB, Sajesh VK (2014) Growth and Instability of Onion Production in Maharashtra, India. J Glob Econ.
[7] S. Ramesh et al., "Plant Disease Detection Using Machine Learning," 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), Bangalore, India, 2018, pp. 41-45, doi: 10.1109/ICDI3C.2018.00017.
[8] Hiroyuki Takanashi, Hiromitsu Furuya and Seiji Chonan, 16th IEEE International Conference on Control Applications Part of IEEE Multi-conference on Systems and Control Singapore, 1-3 October 2007.
[9] Detection of Basal Stem Rot Disease Using Deep Learning by YU HONG HAW, YAN CHAI Hun, (Senior Member, IEEE), JOON HUANG CHUAH (Senior Member, IEEE), WINGATES VOON, (Member, IEEE), SITI KHAIRUNNIZA-BEJO4, NUR AZUAN HUSIN 4, POR LIP YEE 5, (Senior Member, IEEE), AND KHIN WEE LAI 1, (Senior Member, IEEE).
[10] Analysis of Effectiveness of Augmentation in Plant Disease Prediction using Deep Learning"Proceedings of the Fifth International Conference on Computing Methodologies and Communication (ICCMC 2021) IEEE Xplore Part Number: CFP21K25-ART" Jithy Lijo.
[11] Kathole, A. B., Vhatkar, K. N., Kumbhare, S., Katti, J., & Kimbahune, V. V. (2024). IoT-Based Smart Agriculture for Onion Plant Disease Management: A Comprehensive Approach. International Journal of Intelligent Systems and Applications in Engineering, 12(13s), 472-476.
[12] Detection of onion leaf Disease Using Hybridized Feature Extraction and Approach, Gbadebo, G. O., Alhassan, J. K., & Ojerinde, O. A. (2022). 2022 5th Information Technology for Education and Development (ITED) IEEE.
[13] AGRO-ECOSYSTEM ANALYSIS (AESA) BASED INTEGRATED PEST MANAGEMENT (IPM) https://farmer.gov.in/imagedefault/ipm/onion.pdf
[14] Jahan, Mushrat. (2023). A DEEP LEARNING-BASED ONION LEAF DISEASE CLASSIFICATION. 10.13140/RG.2.2.36220.72329.
[15] Onion Disorder: Fusarium Basal Rot by Karen Delahaut & Walt Stevenson, University of Wisconcin.
[16] World Economic Forum – AI4AI- Artificial Intelligence for Agricultural Innovation. Community Paper March 2021.
[17] DISEASES AND INSECT PESTS OF ONION IN INDIA - https://www.ishs.org/ishs-article/358_43
[18] NCBI - https://www.ncbi.nlm.nih.gov/pmc/articles/PMC9415854/
[19] ICAR- Directorate of Onion & garlic research (https://dogr.icar.gov.in/index.php?option=com_content&view=article&id=121&Itemid=192&lang=en)
[20] Vikaspedia (https://vikaspedia.in/agriculture/crop-production)
[21] Deep learning models for plant disease detection and diagnosis Konstantinos P. Ferentinos Hellenic Agricultural Organization “Demeter”, Institute of Soil & Water Resources, Dept. of Agricultural Engineering, 61 Dimokratias Av., 13561 Athens, Greece. https://doi.org/10.1016/j.compag.2018.01.009
[22] Mohanty, S.P., Hughes, D.P., Salathé, M., 2016. Using deep learning for image-based plant disease detection. Front. Plant Sci. 7http://dx.doi.org/10.3389/fpls.2016.01419. Article: 1419.
[23] Neelakantan. P, Analyzing the best machine learning algorithm for plant disease classification, Materials Today: Proceedings - https://doi.org/10.1016/j.matpr.2021.07.358
[24] Machine Learning for Plant Leaf Disease Detection and Classification – A Review by L. Sherly Puspha Annabel, T. Annapoorani and P. Deepalakshmi are with the of Information Technology, St. Joseph’s College of Engineering, Chennai, Tamil Nadu.
[25] ConvolutioJinal neural Network from Scratch by Luís Fernando Torres https://medium.com/latinxinai/convolutional-neural-network-from-scratch-6b1c856e1c07
[26] Demilie, W.B. Plant disease detection and classification techniques: a comparative study of the performances. J Big Data 11, 5 (2024). https://doi.org/10.1186/s40537-023-00863-9.