IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Potato Plant Disease Detection And Classification Using Machine Learning Algorithm

[1] **Yugal S. Bansod**, [2] **Rohan V. Nandanwar**, [3] **Sumit M. Timande,** [4] **Rituja G. Mudhane,** [5] **Prof. Priyanka Patil**

[1,2,3,4] Student at Prof Ram Meghe College of Engineering and Management Badnera Amravati, India

[5] Assistant Professor at Prof Ram Meghe College of Engineering and Management Badnera Amravati, India

Abstract: Potato crops offer numerous advantages to human life, with one of the most significant being their high carbohydrate content, which serves as a primary source of nutrition for humans. The progress in potato agriculture holds vital importance for sustaining human life. However, various challenges impede the development of potato farming, including diseases that target potato leaves. If left unaddressed, these diseases can lead to decreased production and potential crop failure. Notable among these leaf diseases is early blight, caused by the fungus ‘Alternaria solani’, and late blight, caused by the microbe ‘Phytophthora infestans’. Each disease presents distinct symptoms, enabling farmers to take preventive measures upon their identification. However, due to the specialized knowledge required, accurate and timely disease identification proves challenging for the average farmer. In response to this, technology in informatics, particularly digital image processing, can offer a solution for disease identification in leaves potato. In this research aims to propose an effective method for detecting diseases in potato leaves, utilizing three types of data: healthy leaves, early blight, and late blight. The identification process employs deep learning techniques, specifically utilizing the Convolutional Neural Network (CNN) architecture. The result of this research is that the 70:30 data segmentation produces better accuracy than the 80:20 data segmentation. The accuracy obtained is 98 % on training data and 97 % on validation data using 32 batch sizes at 50 epochs.

Keywords **– CNN (Convolutional Neural Network), late blight, early blight, healthy, identification, leaves, disease, pixel, Epoch, Crops, leaves**

I. INTRODUCTION

Potato plants offer numerous advantages for human well-being, primarily due to their carbohydrate content, addressing a fundamental human dietary requirement. However, farmers encounter challenges in the cultivation of potato crops, particularly in the format of diseases affecting potato leaves. Failure to promptly address these diseases has the potential to result in a reduction in potato farm income, consequently impacting food production negatively Ease of Use. Therefore, timely detection of plant diseases is crucial for effective control and prevention measures. Potato plants are susceptible to specific diseases that target their leaves, such as early blight and late blight. Regions characterized by cold and humid conditions contribute to the prevalence of leaf diseases in potato plants. Potato plants are susceptible to specific diseases that target their leaves, such as early blight and late blight. Regions characterized by cold and humid conditions contribute to the prevalence of leaf diseases in potato plants. **Early blight**, a potato leaf disease, manifests initial symptoms through the medium of circular spots on the middle or edges of the leaves, as depicted in below Fig. 1.1.

Subsequently, these spots enlarge, causing the leaves to turn brown. The fungus 'Alternaria solani' is identified as the causal agent of this leaf disease. **late blight**, another potato leaf disease, is primarily caused by the microbe 'Phytophthora infestans'. Leaves affected by late blight exhibit black lesions, as illustrated in above Fig.1.3, leading to significant damage to the plant. Additionally, **Healthy,** Healthy potato leaves typically exhibit vibrant green coloration, turgid and upright posture, and a lack of visible lesions, spots, or discoloration, as illustrated in above Fig.1.2.

Fig 1.1

Fig 1.2

Fig 1.3

The identification process serves as a valuable tool for agricultural managers, enabling them to efficiently address issues related to unhealthy or abnormal plants. In the agriculture, digital image studies have become prevalent with the advancement of technology. These studies focus on various aspects, including disease identification and the assessment of agricultural production quality. A specific digital image study pertains to the identification of leaf rot in potato plants.

The main goal of this research is to develop a system that assists farmers and agricultural managers in identifying diseases affecting potato leaves, utilizing data derived from potato leaf images. The identification process categorizes potato plants into three distinct groups: those with healthy leaves, late blight, and early blight. To achieve this, the study employs the CNN architecture, a Deep Learning method. The data utilized for this study consists of disease-related information on potato leaf images sourced from the Kaggle website under the name Plant Village.

II. LITERATURE REVIEW

The study utilizes machine learning techniques and specifically employs CNN to distinguish and categorize images associated with these conditions. The findings indicate that CNN performs effectively in categorizing these items, achieving a validation accuracy of 90%. Such endeavors hold significant relevance for the agriculture industry as they contribute to the accurate identification of potato plant issues, facilitating improved management and decision-making processes [01]. Novel model designed to effectively identify and detect diseases in potato leaf stands through image processing. Specifically, the CNN model is employed among various machine learning methods for disease identification in leaf of potato photos. The approach involves implementing a CNN-based Utilizing a sequential model for prediction the occurrence of diseases in potato leaves, achieving a model accuracy of 94.2%. The model undergoes testing on both healthy and diseased potato leaves to differentiate between the two states. Subsequently, the algorithm is applied to the images, classifying the potato tree's leaves as either healthy or unhealthy. [03]. Potato crop losses caused by diseases, particularly blight, impacting the leaves. Blight leads to leaf deterioration, disrupts photosynthesis, and eventually destroys the entire plant, resulting in significant losses for farmers. The integration of Deep Learning and Artificial Intelligence in agriculture provides a promising solution for early blight identification in potato leaves, mitigating potential damages. The paper introduces a trained and validated model using CNN technology to distinguish between healthy potato leaves and those affected by late blight or early blight. Impressively, the model achieves an accuracy of 95.84% based on test dataset used in this study [08].

The crucial role of the agricultural sector in national development, especially in a country like India, where approximately 65% of the population depends on agriculture. The susceptibility of crops to diseases is heightened by diverse seasonal conditions, often starting in the leaves, and spreading throughout the crop, impacting its variety and yield. The multitude of plant diseases presents a significant challenge for human visual identification and categorization. Timely diagnosis is essential as it helps prevent the potential spread of diseases. Consequently, an automated method employing CNN for the rapid and user-friendly identification and categorization of plant leaf diseases is proposed. The research includes a comparative study, contrasting CNN-based disease detection with traditional models, utilizing a manually collected and pre-processed dataset of potato plant leaves [09].

III. PROBLEM DEFINITION

Potato leaf diseases pose a significant threat to agricultural productivity, as they can result a substantial reduction in crop yield. This reduction in yield directly impacts farmers' income also in food security. Potatoes constitute a fundamental food crop for numerous communities, and any decline in their production due to diseases can have far-reaching consequences. Farmers rely heavily on success of their crops for economic sustenance, and the susceptibility of potato plants to diseases adds an extra layer of vulnerability. Addressing and managing potato leaf diseases is not only essential for safeguarding farmers' incomes but is also crucial for ensuring a stable and secure food supply for communities that depend on this vital crop. Therefore, efforts to mitigate and prevent these diseases play pivotal role in sustaining both agricultural livelihoods also in food security.

Potato diseases are among the most devastating plant diseases globally, significantly impacting both the productivity and quality of potato crops. The consequences are extensive, affecting both individual farmers and casting a shadow on the broader agricultural industry. The constant threat of potatoes succumbing to various diseases poses a continuous risk to crop yields, putting food security, economic stability in jeopardy. Consequently, addressing, and mitigating potato diseases becomes a necessity to sustain agricultural livelihoods and ensure a reliable food supply. Researchers, farmers, and policymakers are compelled to explore innovative strategies, technologies, and preventive measures to protect potato crops from the harmful consequences of these pervasive diseases.

Diseases such as late blight and early blight exert a profound influence on both the quality and quantity of potato crops. The manual interpretation of these leaf diseases proves to be a time-consuming and cumbersome task. The intricate nature of identifying and distinguishing between different potato diseases poses a considerable challenge for farmers and agricultural practitioners. As these diseases can significantly impact the overall health and yield of potato plants, there is a growing recognition of the need for efficient and timely detection methods. The advent of technology, particularly through the application of advanced image processing, machine learning techniques, offers promising avenues to streamline and expedite the identification process, providing farmers with quicker and more accurate tools to address these critical issues in potato cultivation.

IV. PROPOSED METHODOLOGY

The paper presents a research framework depicted in Fig 4.1, illustrating stages of research completion. The framework comprises four distinct stages: dataset collection, image data pre-processing, training data, and data evaluation. These stages provide a structured approach to the research process, guiding the progression from collecting datasets to pre-processing image data, training the data, and finally, evaluating the results.

Fig 4.1 Research Framework

Dataset Collection: The dataset employed in that study consists of images depicting the leaves of a potato plant, categorized into three classes: healthy leaves (illustrated in Fig 1.2), early blight (depicted in Fig 1.1), and late blight (also shown in Fig 1.3). This dataset is from the Kaggle website, specifically from the "Plant Village" collection uploaded by ARJUN TEJASWI. The dataset utilized comprises 1000 instances of late blight data, 1000 instances of early blight data, and 152 instances of healthy leaf data. A detailed breakdown of the data is provided in Table 1, outlining the agreed-upon specifications for each class.

Samples	Number	Repository
Late Blight	1000	Kaggle (Plant Village)
Early Blight	1000	
Leaf Healthy	152	
Total	2152	

Table 1 Detail Datasets

All images used in this topic will be resized to 256x256 to speed up processing. Fig 4.2 are examples of pictures from each data class used Fig 4.2 (a) leaf healthy, (b) early blight and (c) late blight.

Fig 4.2

Preprocessing Data: At this stage, there are total 2152 images from three classes (late blight, early blight, and healthy leaf) are utilized, as depicted in Figure 4.2. In Table 2, the distribution details of each class's data are presented, segregating them into training and testing datasets with a division ratio of 80:20 and 70:30, respectively. The subsequent analysis of outcomes from each data split seeks to identify the most effective ratio for partitioning the dataset. To expedite the classification process, the initial leaf images are resized to 256x256. This resizing step is implemented to enhance the efficiency of the classification process while maintaining the essential information from the leaf images.

Datasets	**80:20**		**70:30**	
	Train	Val	Train	Val
Late Blight	**400**	**100**	**350**	**150**
Early Blight	**400**	**100**	**350**	**150**
Leaf Healthy	**122**	**30**	**105**	**45**
Total	**922**	**230**	**805**	**345**

Table 2 Distribution Datasets

The data distribution employed in that study follows a division of 80:20 and 70:30, omitting in use of a 90:10 split. This decision is based in the consideration of the 152 healthy leaf data instances. Allocating only 10% of the data for validation in a 90:10 split would not be sufficient for the validation process, especially given the batch size used on this study, which is set at 32 batches. Hence, the chosen data distribution ratios ensure an adequate representation for both training and validation processes, maintaining the robustness of the model evaluation.

Convolution: The CNN employs filters to identify the features present in the leaf image. During the convolution process, matrix multiplication occurs between the filter and the corresponding leaf image area. As illustrated in Fig. 4.3, the convolution process involves multiplying the pixels in the image by the pixels in the filter. In this study, we will employ a total of 4 convolution layers to improve the network's capacity to extract and identify pertinent features from the input leaf images.

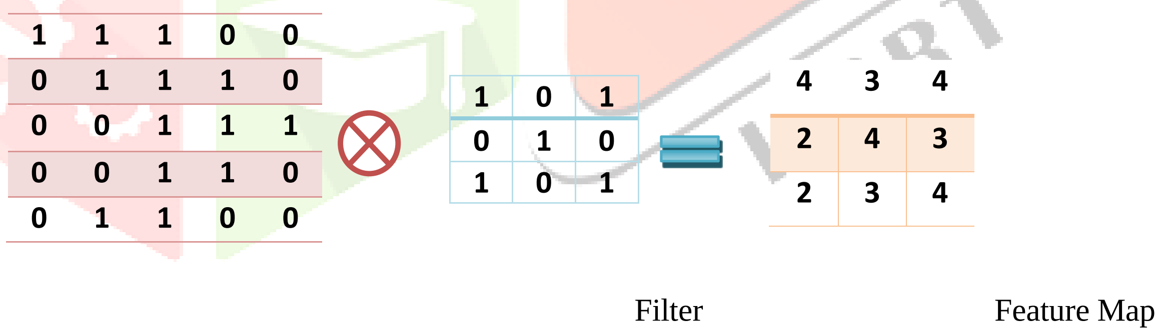


Fig 4.3

Pooling: Following the convolution operation, the subsequent step involves pooling, with MaxPooling being a commonly used technique. Pooling is a process aimed at obtaining images with reduced pixel dimensions while preserving essential information. Depicted in Fig 4.4, the pooling process involves selecting the highest pixel value within a specific pixel area of the image. This constructive process serves to diminish the size of each image, thereby accelerating the classification process. The utilization of pooling contributes to computational efficiency by reducing the computational load while retaining crucial image features.

7	3	5	2
8	7	1	6
4	9	3	9
0	8	4	5

Maxpool

8	6
9	9

Fig 4.4

Classification: The subsequent step involves classifying images using the CNN architecture within the framework of Deep Learning. The CNN architecture, a part of supervised learning, involves training the network with existing image data to identify and categorize specific target variables related to images.

Within the CNN architecture, the convolutional layer play's role in aiding neural networks in recognizing potato leaves based on inherent attributes. By analysing the pixel information within the images, the neural network becomes proficient in identifying features specific to potato leaves.

For this research, images sized 256x256x3 are employed, indicating a 256x256 image with three channels—red, green, and blue (RGB). The initial leaf image undergoes convolution with a filter, followed by MaxPooling to decrease image resolution while preserving quality. Following this, fully connected layer incorporates a flattening process, converting the feature map generated from pooling into a vectorized form. For a more in-depth comprehension, consult the CNN architecture illustrated in Fig. 6.

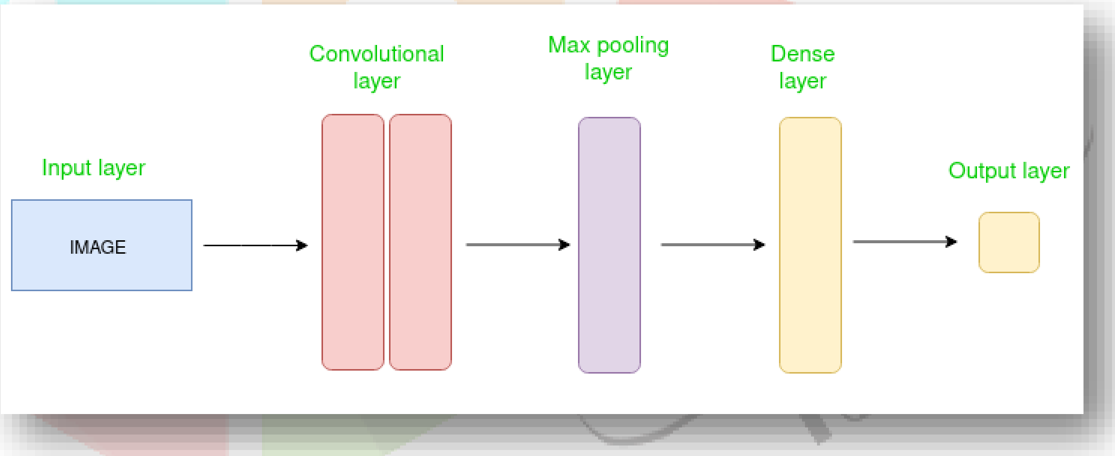


Fig 4.5 CNN Architecture

The research introduces a proposed model for the CNN architecture designed to identify diseases in potato leaves, as outlined in Table 3. This model incorporates 4 convolution layers and 4 MaxPooling layers, showcasing the architectural specifications for effectively recognizing and classifying diseases in potato plants.

Table 3 CNN MODEL

Layer	Output Shape	Param
sequential (Sequential)	(32, 256, 256, 3)	0
conv2d (Conv2D)	(32, 254, 254, 32)	896
max_pooling2d (MaxPooling2D)	(32, 127, 127, 32)	0
conv2d_1 (Conv2D)	(32, 125, 125, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(32, 62, 62, 64)	0
conv2d_2 (Conv2D)	(32, 60, 60, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(32, 30, 30, 64)	0
conv2d_3 (Conv2D)	(32, 28, 28, 64)	36928
max_pooling2d_3 (MaxPooling2D)	(32, 14, 14, 64)	0
conv2d_4 (Conv2D)	(32, 12, 12, 64)	36928
max_pooling2d_4 (MaxPooling2D)	(32, 6, 6, 64)	0
conv2d_5 (Conv2D)	(32, 4, 4, 64)	36928
max_pooling2d_5 (MaxPooling2D)	(32, 2, 2, 64)	0
flatten (Flatten)	(32, 256)	0
dense (Dense)	(32, 64)	16448
dense_1 (Dense)	(32, 3)	195
Total params: 183,747 Trainable params: 183,747 Non-trainable params: 0		

V. RESULTS & EVALUATION

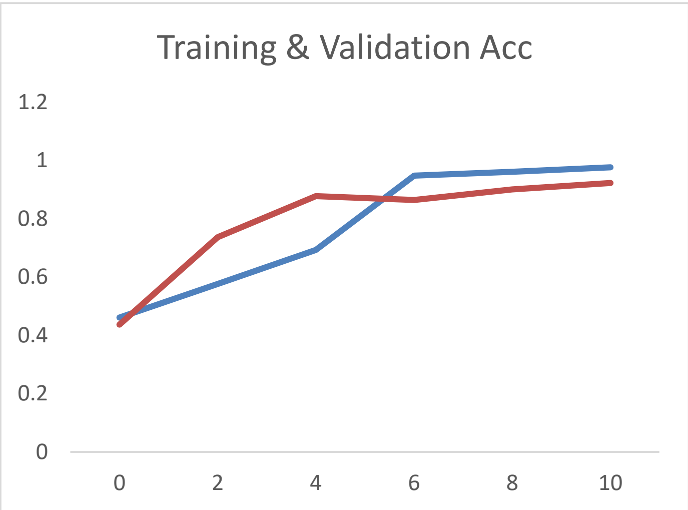
An epoch in neural network training involves completing one round of the entire dataset until it returns to the initial stage. When utilizing a neural network model for training data, employing only one epoch can be burdensome as it may stress the training process due to the substantial amount of data involved. To address this, it becomes necessary to divide the data into batches, known as batch size. In this study, a batch size of 32 was employed, and the researchers adjusted the number of epochs to match the ratio of batch sizes to the total number of samples used.

The next step is to conduct training on the potato leaf image which has been divided by the fit model. Table 4 is the result of the fit model in the 70:30 data division, then Table 5 shows the results of the fit model in the 80:20 data division. It can be seen from epoch 1 to epoch 10 that the accuracy value on the train data and this accuracy value on the validation data has increased.

Epochs	Data Training		Data Testing	
	Acc	Loss	Val Acc	Val Loss
1	0.4611	0.6103	0.4364	0.5946
2	0.5758	0.5156	0.7364	0.3630
3	0.6930	0.3964	0.8773	0.2759
48	0.9478	0.0898	0.8636	0.2510
49	0.9605	0.0666	0.9000	0.1728
50	0.9758	0.0560	0.9227	0.1207

Table 4 Result from Fit model 70:30

Table 4 displays the classification outcomes for the training data in the 70:30 split. In the initial epoch, the accuracy on the training data is documented at 46%, along with a corresponding loss value of 61%. As the epochs advance, significant enhancements are evident, and by the 50th epoch, the accuracy achieves an impressive 97%, accompanied by a minimal loss value of 0.05%. Conversely, on the validation data, the accuracy commences at 43% in the first epoch, with a loss value of 59%. Over subsequent epochs, the accuracy consistently improves, reaching 92% by the 10th epoch, while the loss value decreases to 12%. These findings underscore the model's proficiency in learning and generalizing effectively from the training data to the validation data across multiple epochs.



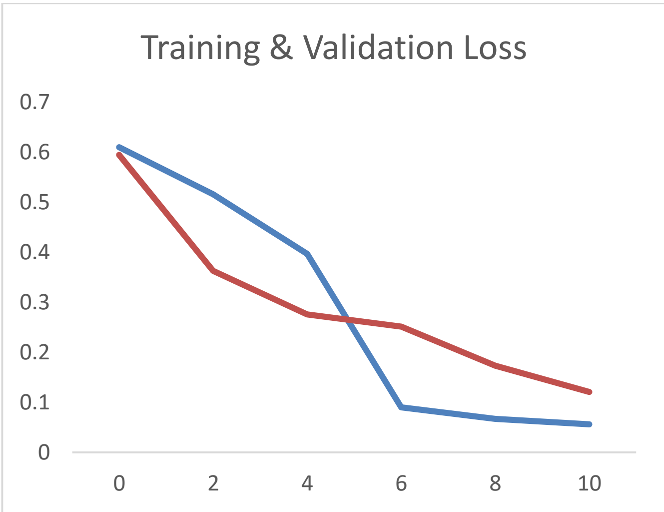


Fig. 4.6 Training and Validation from 70:30 data dividing a) Accuracy b) Loss.

Fig 4.6 (a) illustrates the accuracy graph, while Fig 4.6 (b) depicts the loss graph for the 70:30 data division. The blue line corresponds to training data, and the orange line represents the validation data. These graphs collectively indicate effectiveness of the model, as the consistent and stable increase in accuracy and simultaneous decrease in loss across each epoch reflect the robustness of the fit model. The observed patterns underscore the reliability in the proposed approach and its capability to learn and generalize effectively.

Epochs	Data Training		Data Testing	
	Acc	Loss	Val Acc	Val Loss
1	0.4412	0.6077	0.4364	0.5796
2	0.6522	0.4482	0.7909	0.3902
3	0.8005	0.3083	0.8091	0.3253
48	0.9258	0.1222	0.9182	0.1352
49	0.9629	0.0610	0.9636	0.0861
50	0.9655	0.0623	0.9273	0.1755

Table 5 Result from Fit model 80:20

Table 5 presents the outcomes when the data is split into 80:20. In the first round, the model achieved a 44% accuracy on the training data, with a loss of 60%. As the training continued through 50 epochs, the accuracy increased steadily, reaching 96% with a low loss of 0.06%. Similarly, for the testing data, the initial accuracy was 43% with a loss of 57%, improving over epochs to reach an accuracy of 92% and a loss of 17% by the 50th epoch. These results indicate how well the model learned from the train data and performed on unseen testing data.

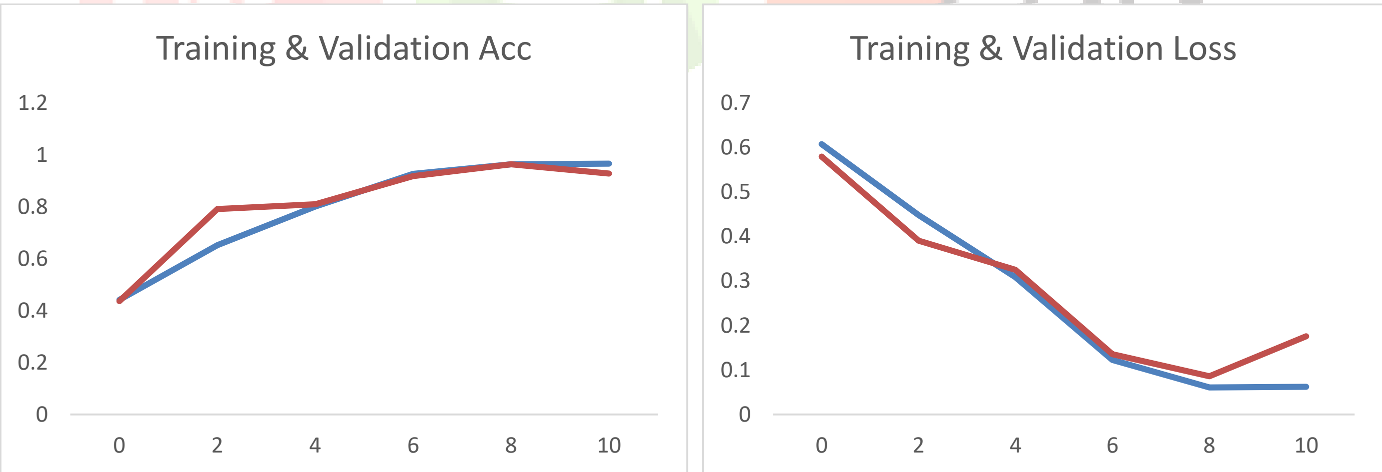


Fig 4.8 Training and Validation from 80:20
data dividing a) Accuracy b) Loss.

In Fig 4.8 (a), you can see a graph showing how accurate the model is, and in Fig 4.8 (b), there's a graph showing how much the model is learning or forgetting, called loss. These graphs confirm that the model created with an 80:20 split on the data is good because the accuracy increases, and the loss decreases consistently for each attempt.

Summing up the study, it seems that the dataset's distribution doesn't have a significant impact on the accuracy results. The tests on potato leaf data slightly fevered the 70:30 split compared to the 80:20 split. The images were resized to 256x256, and the classification was performed over 50 rounds, each using group of 32 images, with a total of 2152 images. For the 70:30 split, the training data had 97% accuracy, and the validation accuracy was 92%. Meanwhile, for the 80:20 split, the training data accuracy was 96%, and the validation accuracy was 92%.

VI. CONCLUSION

This Our work suggests that how we divide the dataset has an impact on how accurate the model becomes. In the case of potato leaf data, dividing it into 70% for training and 30% for testing yielded good results. The images were resized to 256x256. By the 50th attempt, with a batch size of 32 images, the model attained a 97% accuracy rate on the training data and achieved a 92% accuracy rate on the validation data.

VII. FUTURE WORK

The likelihood of potato illnesses occurring is difficult to forecast. Based in the findings of our study, we will implement a system that will assess whether a plant is healthy or not. The project is composed of hardware and software. We will consequently utilize mix hardware and software in the future to obtain better and real-time output.

VIII. REFERENCES

[1] Riya Verma, "CNN based Leaves Disease Detection in Potato Plant", 2023 6th International Conference on Information Systems and Computer Networks (ISCON) | 979-8-3503-4690 696-1/23/$31.00 ©20 ©2023 IEEE | DOI: 10.1109/ISCON57294 2023.10112080

[2] Abhinav Baranwal, Mehul Mishra, Aryan Goyal, Yogesh, "Potato Plant Disease Classification Through Deep Learning", 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON) | 978-1-6654-9602-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/COM-IT-CON54601.2022.9850701

[3] Pankaj Kumar Shukla, Sathiya S, "Early Detection of Potato Leaf Diseases using Convolutional Neural Network with Web Application", 2022 IEEE World Conference on Applied Intelligence and Computing (AIC) | 978-1-6654-7988-2/22/$31.00 2022 IEEE | DOI: 10.1109/AIC55036.2022 9848975.

[4] Choyon Chandra Bonik, Flora Akter, Md. Harunur Rashid, "A Convolutional Neural Network Based Potato Leaf Diseases Detection Using Sequential Model", 2023 International Conference for Advancement in Technology (ICONAT) | 978-1-6654-7517-4/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICONAT57137.2023.10080063

[5] Md Rahmatul Kabir Rasel Sarker, Nasrin Akter Borsha, Md. Sefatullah, "A Deep Transfer Learning-Based Approach to Detect Potato Leaf Disease at an Earlier Stage", 2022 Second International Conference on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT) | 978-1-6654-1120-2/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICAECT54875.2022.9807963

[6] Muh. Syahabuddin Hylmi, Wiharto, Esti Suryani, "Detection of Potato Leaf Disease Using Multi- Class Support Vector Machine Based on Texture, Color, and Shape Features", 2022 International Conference on Electrical and Information Technology (IEIT) | 978-1-6654-5303-5/22/$31.00 ©2022 IEEE | DOI: 10.1109/IEIT56384.2022.9967866

[7] S. Durai[1], T. Sujithra" and M. Mohamed Iqbal 2, "Image Classification for Potato Plant Leaf Disease Detection using Deep Learning", 2023 International Conference on Sustainable Computing and Smart Systems (ICSCSS) | 979-8-3503-3360-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSCSS57650.2023.10169446

[8] Abhinav Baranwal, Mehul Mishra, Aryan Goyal, Yogesh, "Potato Plant Disease Classification Through Deep Learning", 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON) | 978-1-6654-9602-522-531.00 C2022 IEEE DOL 10 1109 COM-IT-CON54601 2022 9850701

[9] Chaturya T, Swathi Y, Vinodkumar, kathik. P, "Detection of Plant Disease using Convolutional Neural Networks (CNN)", 2023 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-9199-0/23/S31.00 2023 IEEE DOI: 10.1109/ICSCDS56580.2023.10105099