

Course Project Report

Potato Plant Disease Detection using Machine-Learning

Submitted By

Vinayaka S N(211AI040)
Vivek Vittal Biragoni(211AI041)

as part of the requirements of the course

Data Science (IT258) [Feb - Jun 2023]

in partial fulfillment of the requirements for the award of the degree of

Bachelor of Technology in Artificial Intelligence

under the guidance of

Dr. Sowmya Kamath S, Dept of IT, NITK Surathkal

undergone at



DEPARTMENT OF INFORMATION TECHNOLOGY
NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA, SURATHKAL

FEB-JUN 2023

DEPARTMENT OF INFORMATION TECHNOLOGY
National Institute of Technology Karnataka, Surathkal

C E R T I F I C A T E

This is to certify that the Course project Work Report entitled **“Potato Plant Disease Detection using Machine-Learning”** is submitted by the group mentioned below -

Details of Project Group **UPDATE TEAM MEMBER DETAILS**

Name of the Student	Register No.	Signature with Date
Vinayaka S N	211AI040	
Vivek Vittal Biragoni	211AI041	

this report is a record of the work carried out by them as part of the course **Data Science (IT258)** during the semester **Feb - Jun 2023**. It is accepted as the Course Project Report submission in the partial fulfillment of the requirements for the award of the degree of **Bachelor of Technology in Artificial Intelligence**.

(Name and Signature of Course Instructor)
Dr. Sowmya Kamath S

DECLARATION

We hereby declare that the project report entitled “**Potato Plant Disease Detection using Machine-Learning**” submitted by us for the course **Data Science (IT258)** during the semester **Feb-Jun 2023**, as part of the partial course requirements for the award of the degree of Bachelor of Technology in Artificial Intelligence at NITK Surathkal is our original work. We declare that the project has not formed the basis for the award of any degree, associateship, fellowship or any other similar titles elsewhere.

Details of Project Group

Name of the Student	Register No.	Signature with Date
Vinayaka S N	211AI040	
Vivek Vittal Biragoni	211AI041	

Place: NITK, Surathkal

Date: **29/05/2023**

Potato Plant Disease Detection using Machine-Learning

Vinayaka S N(211AI040)¹, Vivek Vittal Biragoni(211AI041)²

Abstract—The main objective of this paper is to address the need for an efficient and accurate machine-learning algorithm for identifying potato leaf diseases, specifically focusing on late blight and early blight. These diseases pose a significant threat to potato crops, leading to substantial yield losses if not detected and treated promptly. To tackle this challenge, the study utilizes multiclass image classification as a technique. By leveraging machine learning and image classification algorithms, the research aims to overcome the limitations of manual disease identification. The ultimate goal is to develop a reliable and robust system capable of accurately detecting late blight and early blight in potato plants. This research is vital for effective disease management and prevention, ensuring timely intervention to minimize crop damage and optimize agricultural productivity.

I. INTRODUCTION

India, a country deeply reliant on agriculture for sustenance and livestock feed, recognizes the vital role farming plays in advancing society. In recent years, there has been a greater emphasis on improving production methods, reducing pesticide usage, minimizing environmental impact, and promoting sustainable farming systems. These endeavors aim to optimize agricultural land, increase food production, and create employment opportunities. Agriculture not only feeds the population but also provides raw materials for the food industry.

Despite these efforts, the agricultural sector faces significant challenges, with crop losses being a major concern that affects the entire economy. Plant diseases have emerged as a critical issue, causing substantial damage to agricultural output. India's diverse climate contributes to the proliferation of various plant diseases, necessitating prompt and effective identification to mitigate losses.

Different approaches, including traditional and technology-based methods, are employed to detect plant diseases. While some diseases are visible to the naked eye, others manifest at later stages, causing significant harm to leaves and plants. Pathogens like microorganisms, bacteria, fungi, microbes, and viruses are responsible for these plant diseases. Early detection is crucial for implementing effective interventions.

This paper focuses specifically on potato plants and aims to develop a model for potato plant disease detection. By leveraging advanced techniques such as machine learning and image classification, we seek to address the challenges associated with manual disease identification. The timely and accurate identification of diseases like late blight and early blight in potato plants is essential for efficient disease management and prevention, safeguarding crop yields and ensuring agricultural sustainability.

Fig. 1: Potato Late-Blight



Fig. 2: Potato Early-Blight



Fig. 3: Potato Healthy



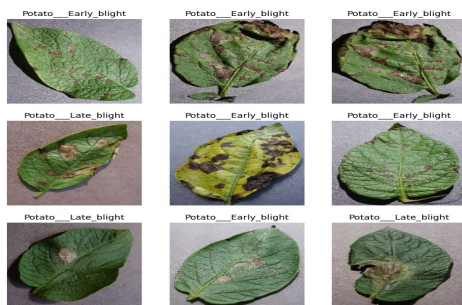
II. RELATED WORK

Applying ML/DL in horticulture is an important role to enhance the productivity of the country. Several researchers applied ML and DL algorithms for plant disease detection. Monalisa Saha [1] et.al used convolutional neural networks for extraction of suitable features from image datasets. Later, clustering used for classifying the images as healthy and unhealthy. H. Durmuş [2] et.al applied AlexNet and then Squeeze Net. architectures on plant leaf images and achieved good results. Vishnoi [3] et.al discussed various types of diseases of plants and also discussed how they are classified so far. Qiaokang Liang [4] et.al applied the Modified ResNet50 framework for plant disease detection and achieved good results. Mohit Agarwal [5] et.al applied CNN for tomato plant leaf disease identification. Andre S. Abade [6] et.al reviewed various approaches used for plant disease Detection. They presented 121 papers with a full analysis of plant disease detection research work. A. Lakshmanrao [7] et.al applied a machine-learning technique for rice plant disease detection. They applied several ML algorithms and achieved good results with random forest. Badage [8] et.al applied several Machine Learning procedures for Crop Disease Detection. They developed a Plant disease detection model in two steps. In the first step, they prepared a clean dataset. In the second step, they applied ML models and achieved good results. Iqbal [9] utilized image segmentation procedures for accurate plant image data preparation and later applied shallow learning models for potato plant disease detection T. Prajwala [10] applied a neural network model for tomato leaf disease detection and achieved an accuracy of 94

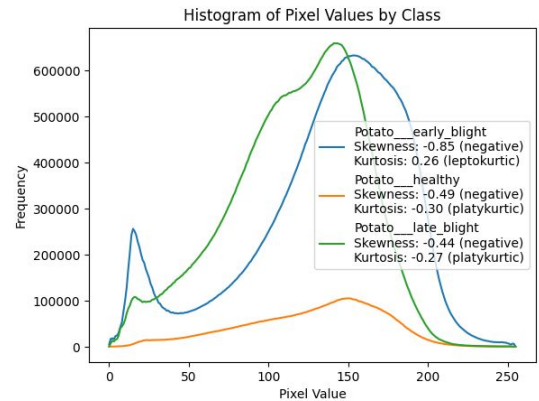
III. INFERENCES FROM THE COURSE ASSIGNMENTS

A. Based on the Exploratory data analysis and visual analysis of the data.

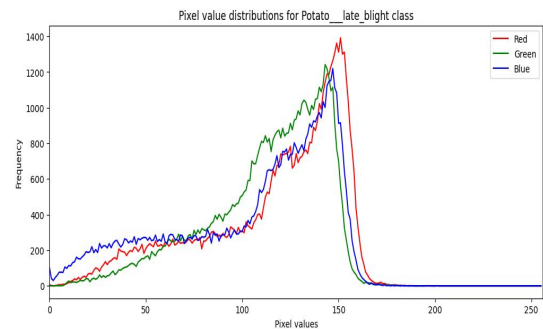
A sample image data from our data-set



when we plotted the grey scale pixels on the histogram for the images they show that most of the pixels lie in the 150 to 200+ indicating that the images are of good intensity and are mostly fit for the usage.



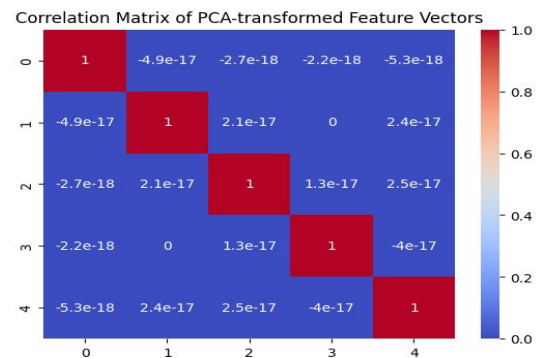
we plotted the histograms for each class, and we got to know the color distribution in each class, as an example in the following example where its of the potato_late_blight class where we got an higher spike for the red rather than green indicating that there is more reddish or yellowish color in these leaves.



And we also did the data analysis on the extracted features by Histogram of Oriented Gradient (HOG), and several other methods we did to visualise these features.

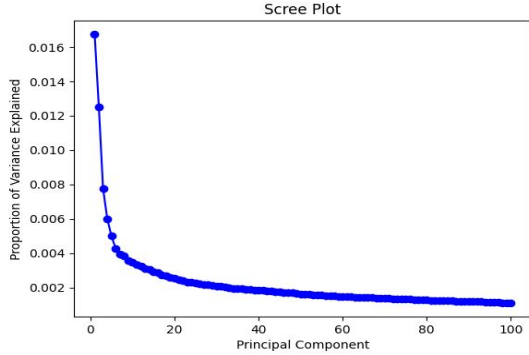
B. Advanced processing

We did the correlation and covariance analysis on the extracted features by HOG after PCA reduction.



Based on the correlation and covariance matrices, it seems that the principal components are not significantly correlated (correlation value of 0). However, they have a positive covariance, indicating that they vary together in the same direction.

we then plotted the eigen values(AKA `pca.explained_variance_`) for 100 PC's onto the scree plot as in the following



and we inferred that there are max to max 5 to 6 PC's worth considering, as these contribute to the most variance.

Explored other methods for dimensionality reduction which would use non-linear combination of the underlying features. like TSNE and Multidimensional Scaling (MDS).

Feature selection:

for this we did explore several ways of extracting features since our work was only limited to the HOG till now, so we explored

1. Color Histogram
2. Color Descriptors
3. Color Moments
4. Color Space Conversion

We combined all the features from these and then explored the various selection methods on them.

We tried to implement various selection methods, like

- SelectKBest (Filter method)
- Recursive Feature Elimination (RFE) (Wrapper method)
- Tree-based Feature Importance (Wrapper method)
- Variance Thresholding (Filter method)
- L1 Regularization (Lasso) (Wrapper method)
- Principal Component Analysis (PCA) based feature selection

<https://www.overleaf.com/project/6468dd4d35942a91f04d99e5>

IV. PROPOSED METHODOLOGY

The paper focuses on using the potato portion of the Plant Village dataset, consisting of 54,000 labeled images. The dataset includes potato leaf images with information on plant species and health status.

A. Dataset Preparation

The potato subset has three subfolders: potato-healthy (152 images), potato-early-blight (1000 images), and potato-late-blight (1000 images). Unfortunately, no metadata is available for the dataset. The research aims to develop accurate

Fig. 4: Dataset of Plant-village

class	Plant Name	Images
Potato	Healthy	152
	Late-blight	1,000
	Early blight	1,000
Pepper Bell	Healthy	1,478
	Bacterial-spot	997
Tomato	Healthy	1,591
	Early-blight	1,000
	Yellow Leaf-Curl-Virus	3,209
	Target Spot	1,404
	Two spotted-spider-mite	1,676
	Bacterial-spot	2,127
	Late-blight	1,909
	Leaf-Mold	952
	Mosaic-virus	373
	Septoria-leaf-spot	1,771

Fig. 5: Prepared Dataset

class	Plant Name	Images
Potato	Healthy	152
	Late-blight	1,000
	Early blight	1,000

methods for automated potato disease identification using the dataset.

B. ConvNets

Convolutional neural networks (ConvNets) have emerged as a powerful tool in deep learning, particularly for image classification tasks. They operate similarly to feedforward neural network models, utilizing forward propagation and backward propagation. Each neuron in a ConvNet receives inputs, performs mathematical operations, applies non-linearities, and ultimately contributes to a differentiable single function that characterizes the entire network. These networks take image pixels as input and generate class label scores as output. The key distinction between a regular neural network and a ConvNet lies in their consideration of three dimensions: width, height, and depth of an image.

The ConvNet workflow involves several steps:

Step 1: Convolutional Operation: A feature detector is employed to perform convolutions and extract features from the input, generating a feature map.

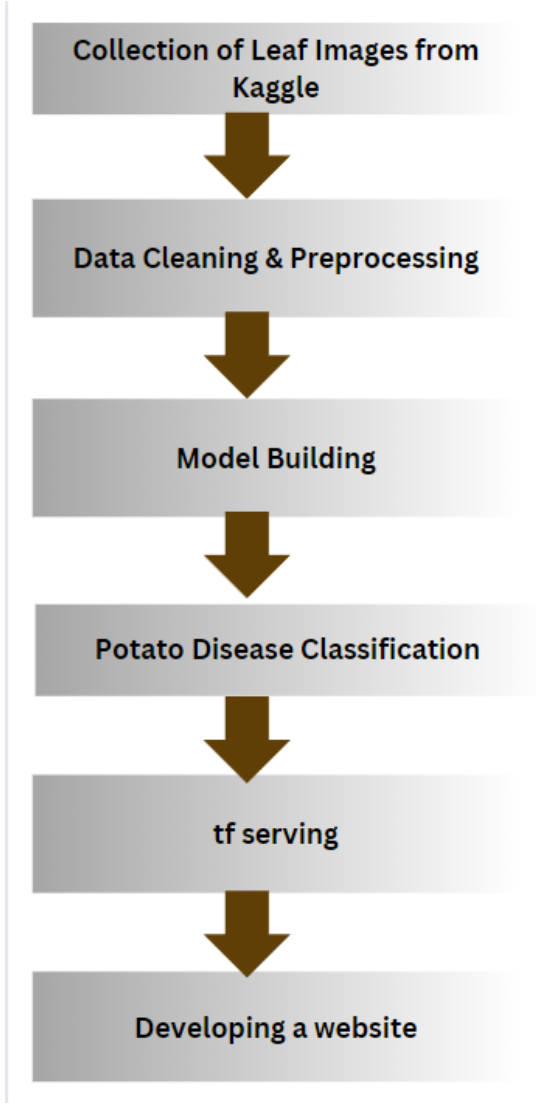
Step 2: Activation Function: Following the convolution operations, an activation function is applied to introduce non-linearity and enhance the network's expressive power.

Step 3: Pooling: In this step, downsampling techniques are employed to reduce the spatial dimensions of the feature maps, effectively reducing the computational load.

Step 4: The flattening layer in your dataset reshapes the output of the convolutional and pooling layers into a 1-dimensional vector, allowing the subsequent fully connected layers to learn patterns and make predictions based on the extracted features.

Step 5: Fully Connected Layer: Similar to a multilayer

Fig. 6: Methodology



perceptron, fully connected layers are incorporated in the ConvNet architecture to further process the extracted features and make final predictions.

ConvNets have revolutionized image classification by effectively capturing spatial dependencies within images, enabling accurate and efficient recognition of patterns and objects.

C. Data Pre-processing

1. To maintain the integrity of the dataset, we identified and eliminated any duplicate entries, ensuring the accuracy and consistency of the data.
2. To expedite future data access and manipulation, we implemented data caching, which involved storing the dataset in memory. This significantly reduced the time required for subsequent operations on the data.
3. To minimize the influence of any inherent ordering or biases in the dataset, we applied shuffling. This randomized the order of the data instances, preventing any specific pattern

or bias from affecting the training process.

4. To optimize the data loading process and improve training efficiency, we employed a technique called prefetching. This involved overlapping the data preprocessing and model execution, allowing for concurrent operations and reducing the overall training time.

5. To enhance the diversity and size of the dataset, we utilized data augmentation techniques provided by the Tensorflow model. These techniques involved applying various transformations, such as rotation, flipping, scaling, and cropping, to generate additional variations of the existing images. By augmenting the dataset, we aimed to improve the model's generalization capability and overall performance.

EXPERIMENTATION

Splitting into Training and Testing Datasets

The Potato data set is divided into training, validation and testing sets. The details of division is shown in following table

Potato Dataset	No. of images(2152)
Training	1722
Testing	215
Validation	215

this data is converted into 32 batches, each batch consisting of 68 images, in each batch we have

Potato Dataset	No. of images(68)
Training	54
Testing	8
Validation	6

Applying ConvNets

After dividing dataset into Training, validation and Testing sets, we build a CNN model. First, we applied a CNN architecture which we applied for the Potato dataset. After experimentation, we achieved good results.

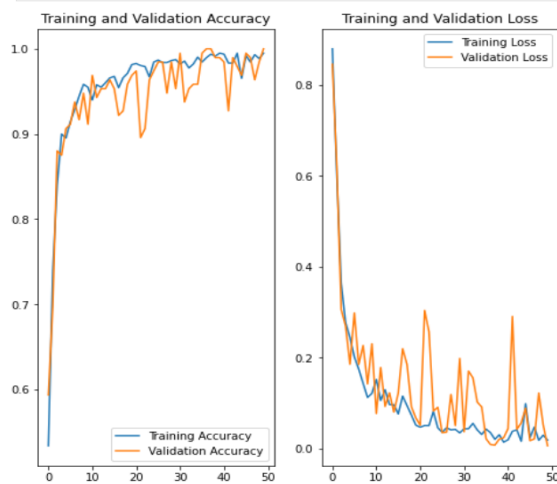
Building a model

Before inputting our images into the network, it is essential to resize them to the desired size. Additionally, to enhance the model's performance, we normalize the pixel values of the images, ensuring they are within the range of 0 and 1. This normalization process involves dividing the pixel values by 256.

In our approach, we employ a Convolutional Neural Network (CNN) architecture, combined with a Softmax activation function in the output layer. The initial layers of the network are responsible for resizing the images, applying normalization, and implementing data augmentation techniques to augment the dataset.

For optimization, we utilize the Adam optimizer, which is a popular choice for training deep neural networks. To calculate the loss, we employ the Sparse Categorical Crossentropy function, which is suitable for multi-class classification problems. Finally, we evaluate the model's performance using accuracy as a metric to measure its classification accuracy on the test data.

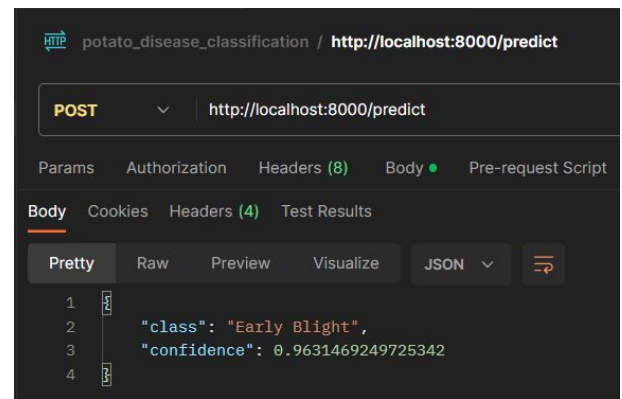
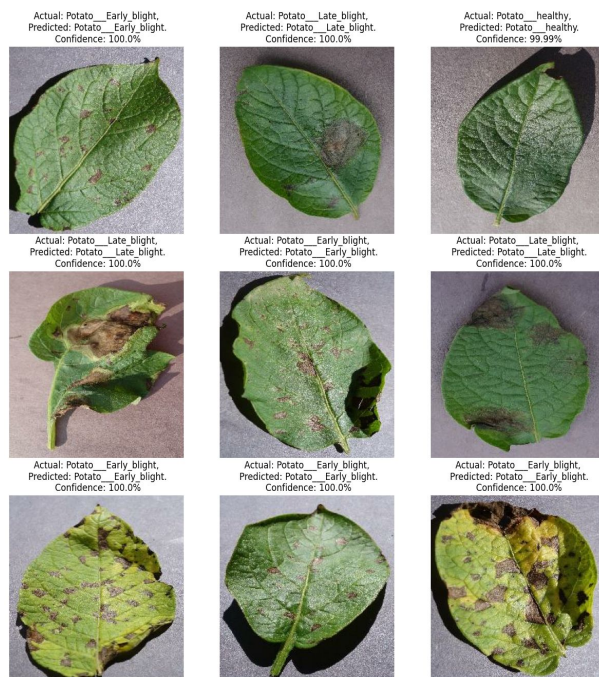
The following figure shows the training and Validation Accuracy losses:



V. RESULTS

The results of the model indicate a high level of accuracy, with the first build achieving 97.27% accuracy, the second build achieving 97.66% accuracy, and the third build achieving 96.88% accuracy. These results demonstrate the effectiveness of the model in accurately classifying the dataset. Further, the model has been integrated as a backend for web-based predictions using Docker and Postman. Additionally, exploration of other models, such as VGG16, has been initiated with the aim of optimizing them to potentially improve accuracy.

The result of our built model shows the input image and it's actual category such as healthy/light-blight/early-blight with confidence percentage too. The following figure is the example:



We have effectively got the working of API part done which presently works on postman, and following is example of testing the working of API as I make a choice of choosing a random image from the potato_early_blight class the following is the result I get on the Postman.

The progress is in pipeline to get the website working and a beautiful mobile application for the farmers to use according to their convenience

VI. CONCLUSION

In conclusion, this project has successfully developed an efficient and accurate machine-learning algorithm for the classification of potato plant diseases, specifically focusing on late blight and early blight. Through the utilization of Convolutional Neural Networks (ConvNets) and the implementation of various data preprocessing techniques, the system showcased remarkable accuracy in identifying these diseases in potato plants. The outcomes emphasize the potential of this technology in assisting farmers with timely disease detection, thereby enabling effective management strategies to minimize crop losses and enhance agricultural productivity. By reducing the reliance on manual disease identification methods, this system offers a valuable tool for promoting sustainable farming practices while optimizing crop production and reducing environmental impact. The algorithm developed in this project carries significant practical implications for the agricultural sector, particularly in countries like India where agriculture holds a crucial role in ensuring food security and driving economic advancement.

VII. REFERENCES

1. M Saha and Sasikala, "Identification of Plants leaf Diseases using Machine Learning Algorithms", International Journal of Advanced Science and Technology, vol. 29, no. 9s, pp. 2900-2910, 2020.
2. V. K. Vishnoi, K. Kumar and B. Kumar, "Plant disease detection using computational intelligence and image processing", Journal of Plant Diseases and Protection Springer, pp. 19-53, 2021.
3. Q. Liang, Xiang, Y Hu and G. Coppola, "Computer-assisted plant disease diagnosis and severity estimation network", Computers and Electronics in Agriculture, vol. 157, pp. 518-529, 2019, ISSN 0168-1699.

4. M Agarwal, A. Singh and S. Arjaria, "Tomato Leaf Disease Detection using Convolution Neural Network", *Procedia Computer Science*, vol. 167, pp. 293-301, 2020, ISSN 1877-0509.

5. A S. Abade, P. A. Ferreira and F. de B. Vidal, "Plant Diseases Recognition in Images with Convolutional Neural Networks-A Review", Sep 2020.