Potato Leaf Disease Detection Using Deep Learning

Renugadevi R
Department of Computer Science and Engineering
Vignan's Foundation Science Technology and Research
Guntur, Andhrapradesh, India
renu.rajaram@gmail.com

Maridu Bhargavi
Department of Computer Science and Engineering
Vignan's Foundation Science Technology and Research
Guntur, Andhrapradesh, India
bhargaviformal12@gmail.com

Shaik Bibi Reshma
Department of Computer Science and Engineering
Vignan's Foundation Science Technology and Research
Guntur, Andhrapradesh, India
201fa04246@gmail.com

Chegu Manasa
Department of Computer Science and Engineering
Vignan's Foundation Science Technology and Research
Guntur, Andhrapradesh, India
chegumanasa1@gmail.com

Jarugula Hari Chandana
Department of Computer Science and Engineering
Vignan's Foundation Science Technology and Research
Guntur, Andhrapradesh, India
harichandana3110@gmail.com

2023 9th International Conference on Smart Structures and Systems (ICSSS) | 979-8-3503-8420-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICSSS58085.2023.10407085

***Abstract*-This study suggests a novel approach for detecting and classifying plant leaf diseases using convolutional neural networks (CNNs), which have shown promise in various image identification applications, including plant disease detection. The method involves utilizing a dataset comprising high-resolution images of both healthy and diseased plant leaves. These images undergo pre-processing and are fed into a well-designed CNN architecture, trained to recognize complex patterns associated with different plant diseases. Transfer learning techniques, utilizing pre-trained CNN models, are explored to enhance classification accuracy and save training time. Additional data augmentation methods, such as rotation and scaling, are applied to improve the model's ability to generalize. The trained CNN model is rigorously calculated on a separate test dataset, demonstrating high accuracy, precision, recall, and F1-score. The results indicate the method's utility in identifying and categorizing plant leaf diseases, showing resilience to environmental factors and leaf orientations. This approach holds promise for early disease identification, enabling prompt responses, and supporting sustainable agriculture for farmers and agricultural specialists.**

Keywords: Plant disease detection, CNN, Deep learning, Image classification, Precision agriculture.

I. INTRODUCTION

This introduction highlights the significance of potatoes as a vital global crop but also emphasizes the challenges posed by diseases, which can severely impact crop output and quality. Traditional disease detection methods relying on human visual examination are subjective and error-prone, leading to the need for automated systems based on computer vision and deep learning(DL), particularly CNN. The study aims to employ CNN technology to develop a reliable system for accurately diagnosing various potato leaf diseases. Early disease identification is crucial for sustainable farming, reducing pesticide use, and ensuring healthy crops. By providing farmers with timely information and targeted interventions, the study intends to enhance agricultural resistance to diseases, increase crop yield, and maximize resource utilization. The research's goal is to create a robust DL model trained on a diverse dataset of potato leaf images, both healthy and diseased, with the potential to revolutionize disease diagnostics in potato farming, improve precision farming methods, and enhance global food security. The goal is to create a trustworthy and effective instrument that will change disease diagnostics in potato farming through testing and validation. In the field of potato disease detection, advanced algorithms leveraging machine learning techniques have been explored to accurately identify and categorize diseases in potato crops [1-3,7]. The importance of this work is to study the growth of plants necessitates observing their biochemical and biophysical characteristics.

II.BACKGROUND OF THE RESEARCH

In the field of potato disease detection, advanced algorithms leveraging machine learning(ML) techniques have been explored to accurately identify and categorize diseases in potato crops. The K-means clustering algorithm, which categorizes objects based on defining features, has been applied to partition potato leaf images into distinct clusters, aiding in disease identification. By segmenting leaves, the algorithm enables precise examination of different regions, facilitating the classification of disease and non-disease areas [3-5]. The study focuses on integrating Artificial Neural Networks (ANN), inspired by biological nerve systems, to differentiate between healthy and diseased potato leaves. Through training on a dataset of potato leaf images, the ANN becomes proficient in recognizing leaf conditions. Combining the ANN's capabilities with K-means clustering offers an auspicious approach to enrich the accuracy and consistency of potato disease identification in agricultural contexts.

Comparison of existing model performance

Table1 lists the different algorithms [5-7] that have been used to identify potato disease and the level of accuracy attained by each algorithm.

Table 1: Performance Comparison of ML Algorithms

ALGORITHM	REF.NO	ACCURACY
ANN	[4]	84-92%
NN	[5]	92%
BPNN	[6]	92%
SVM	[1]	96.83%
KNN	[1]	94.00%
Naïve Bayes	[1]	88.67%
SSD & RCNN	[2]	94.60%

Plant leaves are a valuable tool for tracking a plant's overall health. The aim of this research is to develop a CNN based system that can identify the type of plant leaf disease infection.

III. METHODOLOGY

We employ DL methods to create a robust model for the finding of potato leaf diseases. The methodology consists of a series of carefully designed steps, from data collection to model deployment, which collectively form the core of our approach to addressing the pressing issue of disease management in potato crops. The systematic and comprehensive methodology we present embodies our commitment to delivering an innovative and effective model that can significantly benefit potato cultivation, contributing to increased yields, reduced economic losses, and food security in the agricultural sector. The methodology can be divided into several key stages:

A. Algorithm: DL Algorithm for Potato Leaf Disease Detection

- Gather real-time pictures of potato plants in both lab and field settings, then compile them into a dataset.
- Separate the dataset of Potato Leaf images into sets intended for testing, validation, and training.

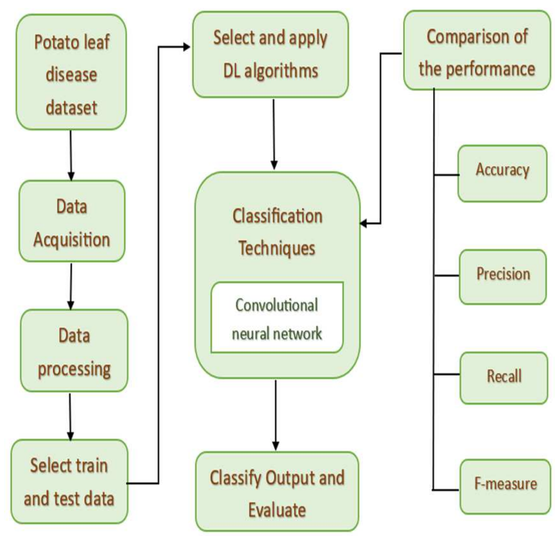

Fig 1. Process involved in disease classification using CNN

- To the interpreted images, preprocessing is applied.
- To the training set, pre-processing is applied.
- Used Google Colab to train and examine the model using the Dataset.
- Potato leaves were mined and segmented to create the Potato Leaf Dataset (PLD) using the annotations from the Dataset and the model's classification output.
- Label the PLD images.
- Apply data augmentation to all images used during the pre-processing stage.
- Assign 80% and 20% of the dataset to training and testing, accordingly.
- Using training data, train the CNN model.
- At the completion of every epoch, use the validation photos to verify the CNN model.
- Keep the Trained Potato Leaf Disease Detection CNN Model Saved.

- Using testing images, testing is conducted to the trained model.

DATASET DESCRIPTION

This study's dataset came from the open Kaggle database, a well-liked source for datasets related to data science and machine learning. In this instance, the dataset was picked on purpose since it is significant to the research on diseases of the potato leaf.

Dataset Contents

The dataset is primarily focused on potato leaf diseases, and it is structured into three distinct categories, each representing a different class or category of plant images. These categories are as follows:

Early Blight

Images of potato leaves infected with the "Early Blight" disease can be found in this category. Early Blight, a fungal disease leading to characteristic symptoms that can be visually identified in images.

Healthy

The "Healthy" category comprises images of healthy potato leaves, free from any disease symptoms or abnormalities. These images serve as a reference for what a normal, unaffected potato leaf should look like.

Late Blight

The "Late Blight" category consists of images of potato leaves afflicted by the "Late Blight" disease. Late Blight is another significant fungal disease that can severely impact potato crops.

Dataset Size

The dataset is relatively modest in size, containing a total of 2,152 images. These images are distributed across the three categories (Early Blight, Healthy, Late Blight) and serve as the primary data for training, validating, and testing the DL model for potato leaf disease prediction.

Table 2: Number of images with diseases

NO. OF IMAGES	DISEASE	TYPE OF DISEASE
1000	Early Blight	Fungal infection
1000	Late Blight	Fungal infection
152	Healthy	Healthy leaves without disease

DATA PRE-PROCESSING

Data pre-processing is a fundamental preparatory phase in the progress of a DL model for potato leaf disease detection. This phase is devoted to refine the dataset, ensuring that it meets the requirements for training a deep neural network effectively. It is a preparatory step to enhance the quality and uniformity of the dataset for deep learning.

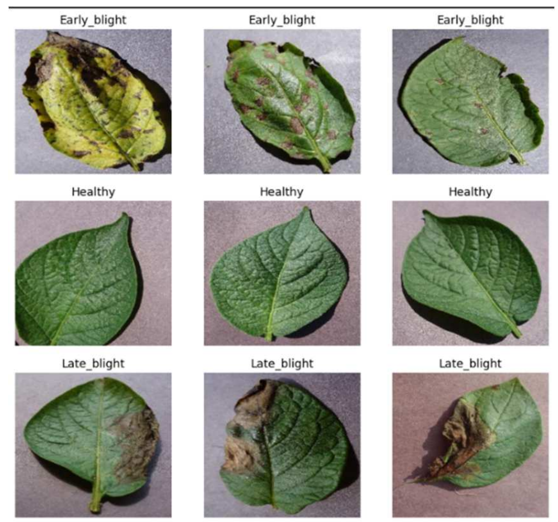

Fig 2. Dataset

We perform operations such as image resizing to a consistent resolution, normalization to scale pixel values, and data augmentation.

Image Resizing

The initial task in data preprocessing is image resizing. This step entails adjusting the dimensions of all the images in the dataset to a uniform and consistent resolution. Common image dimensions are 224x224 or 256x256 pixels. This standardization is essential for deep learning models that require fixed input sizes.

Image Normalization

Normalization is a critical operation within data preprocessing. It includes rescaling the pixel values within the images to a predefined range, typically within the bounds of 0 to 1 or -1 to 1. Here, you scale pixel values to a range of [0, 1] by dividing them by 255 or to a range of [-1, 1] by subtracting the mean and dividing by the standard deviation. Normalization helps with model training by ensuring that the values are centre around zero.

Data Augmentation

Data augmentation involves introducing controlled variations into the dataset, such as rotations, flips, and brightness adjustments. By using these methods, the model can be strengthened against overfitting and become more resilient to changes in the real world. Data augmentation supports the model generalize better to different lighting, angles, and disease variations.

Training and Testing

Training and testing portions comprised the whole dataset into 2 parts. The model was trained using the training dataset, and model performance was assessed using the validation and test datasets. As a result, we divided the datasets for training and testing into 80% and 20%, respectively. 2152 photos were utilized in the dataset for testing, validation, and training. Several data augmentation methods, such as channel shift, brightness, rescaling, rotation, width shift, height shift, shear range, brightness etc., and fill mode nearest to boost the dataset's variety, were functional to the training set. It would solve the overfitting issue and guarantee the model's generalizability.

Convolutional Neural Network (CNN)

CNN is a supervised learning method that targets picture variables and trains on pre-existing datasets to identify images. Based on their characteristics, potato leaves are recognized by the neural network because of the convolutional layer of CNN. A neural network recognizes photos of potato leaves by analysing the pixels in the image. Convolutional neural networks are the fundamental deep learning tool utilized in this work (CNNs). It is one of the most effective methods for simulating intricate processes and recognizing patterns in vast amounts of data, such as picture recognition, is the use of CNNs. The capability of CNN to automatically mine features, particularly from databases containing speech, picture, and video data, is what promotes its use. The division of skill is a crucial and fascinating aspect of the image, albeit it is not always required. When DL is used in the agricultural area for discovery and classification, better results are obtained than with other models. Despite the excellent results that have been reported, there is a lack of diversity in the datasets that were employed. CNNs must be trained on big datasets, each containing thousands of images. CNNs are a subset of convolutional neural networks family that are used for deep learning. CNNs represent an important progression in image recognition classification. They are regularly working in visual imagery analysis as well as picture classification. 2-D neural-networks that obtain input in two dimensions are known as convolutional neural networks. Neural networks make sense for imagine pre-processing since images can be represented in similar ways. Convolution is unique in that as it results in distinguishable features in images. It screens the input before creating a feature map that lists all of the features that have been noticed. Throughout training, these networks pick up the filters in the context of a specific prediction task. Figure 3 illustrates CNN's architecture.

There are layers in CNN which are

- Convolution-layer
- Pooling-layer
- Fully-connected layer

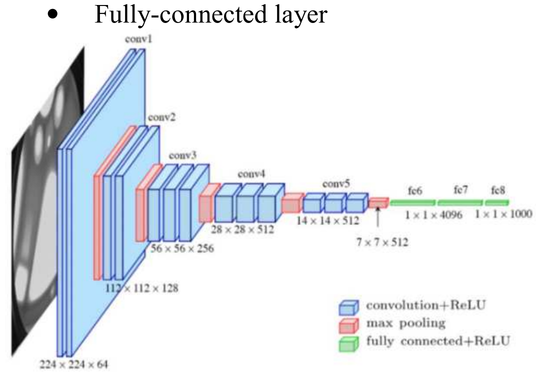

Fig 3. Architecture of CNN

2-D neural networks that obtain input in 2-D are known as CNN. Neural networks make sense for imagine pre-processing since images can be represented in similar ways. Convolution is unique in that it results in distinguishable features in images.

It filters the input before creating a feature map that lists all of the features that have been noticed. Throughout training, these networks pick up the filters in the context of an exact expectation task.

IV EVALUATION METRICS

Accuracy

It is calculated by the quantity of correct predictions divided by the total quantity of predictions (True positive, True Negative, False Positive, False Negative).

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

Precision

Precision is a measure that counts the number of true positive predictions made. It measures the model's accuracy for the minority class.

$$Precision = \frac{TP}{TP + FP}$$

Recall

Division of input positive samples from a class correctly predicted by the model.

$$Recall = \frac{TP}{TP + FN}$$

F1-Score

It measures a model's accuracy by combining precision and recall of that model.

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

V RESULTS

The results of our project are as follows:

Training and testing:

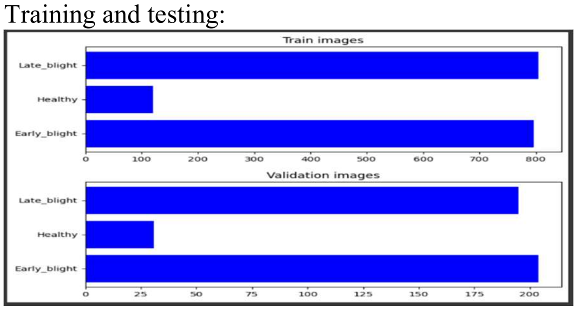

Fig 4. Training and testing number of images

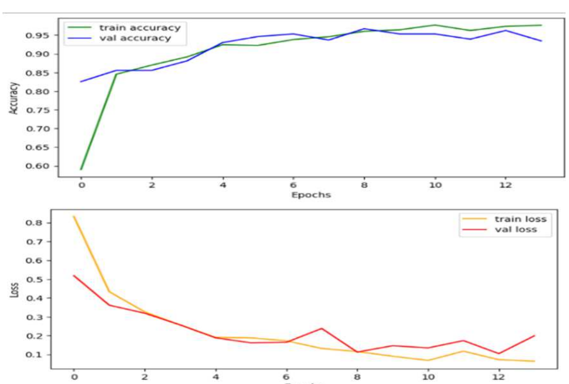

Fig 5. Train and test accuracy

Classification

The prediction on every single image is been done here and we get that the model classifies the data very easily in graphical format. Numerous hidden neurons were tested to identify the maximum accuracy with minimum error possible. The accuracy is shown in Figure 5 &6.

Fig 6. Correct prediction on single healthy image

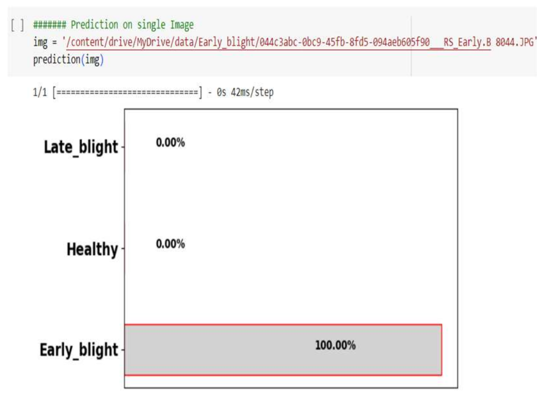

Fig 7. Correct prediction on single early blight image

Fig 8. Correct prediction on single late blight image

When the leaf classification model was put into practice, to identify whether a leaf in the image was healthy or diseased. The model's accuracy can be ascertained by presenting the confusion-matrix and looking at the diagonal values to determine the number of accurate classifications.

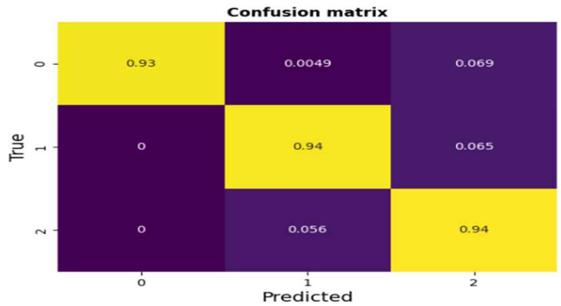

Fig 9. Confusion matrix

Finally, we construct a neural network model, implement it effectively, and obtain desired outcome. The leaf image that is imported into the interface and the message indicating which disease the plant leaf is diseased with are clearly visible in the below fig. 10 and 11.

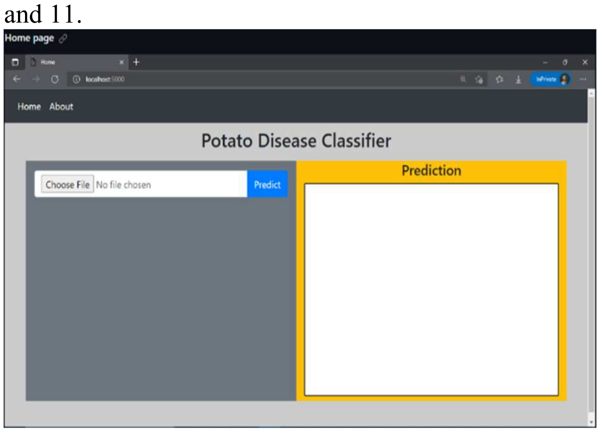

Fig 10. End result of model #Home page

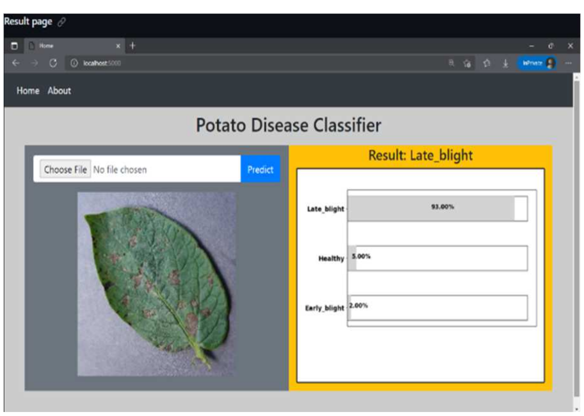

Fig 11. End result of model #Result page

VI CONCLUSION

In conclusion, our work, focused on Potato Leaf Disease Detection using CNN, has demonstrated significant success. By creating a user-friendly interface, employing advanced image processing and deep learning techniques, we have achieved a robust disease classification system. We used CNN to effectively identify and classify healthy and diseased potato leaves. Our results show an impressive classification accuracy, with 97.3% accuracy in distinguishing between healthy and diseased leaves. This achievement not only highlights the potential for early disease detection in the agricultural sector but also underscores the efficiency and accuracy of our approach. Looking forward, this technology has the potential to be extended to the classification of diseases in various plant species, further assisting the agricultural sector, particularly farmers, in a cost-effective and time-efficient manner. This project represents a significant step towards enhancing crop management and agricultural productivity through innovative disease detection methods.

REFERENCES

[1] S Jeyalakshmi and R Radha, "An effective approach to feature extraction for classification of plant diseases using machine learning", 2020.

[2] Viraj A. Gulhane and Ajay A. Gurjar, "Detection of Diseases on Cotton Leaves and Its Possible Diagnosis", 2011.

[3] Bashar Abul, "Survey on evolving deep learning neural network architectures", Journal of Artificial Intelligence, vol. 1, no. 02(2019), pp. 73-82.

[4] Taruna Sharma and Ruchi Mittal, "Classification of Plant leaf diseases: A Deep Learning Method", 2019.

[5] RenugadeviR,Jeya Prakash, SakthivelB,Arockia Raj Y, ‘An Iot-Based System For Effective Covid Patient Health Monitoring With SVM Decision Making’ Turkish Journal of Physiotherapy and Rehabilitation,Vol 32,Issue 32(3) ISSN 2651-4451 | e-ISSN 2651-446X, 2021.

[6] RenugadeviR ,VijaiMeyyappan M, A Promising Method for Predicting Employee Mental Health using Machine Learning Algorithms, Applications of AI in Emerging Research and Education, 2023,pp 279-290,ISBN 978-81-965582-7-7

[7] L. Li, S. Zhang and B. Wang, "Plant Disease Detection and Classification by Deep Learning—A Review," in IEEE Access, vol. 9, pp. 56683-56698, 2021, doi: 10.1109/ACCESS.2021.3069646.