PLANT DISEASES DETECTION & CLASSIFICATION USING DEEP LEARNING TECHNIQUES

Project progress report in partial fulfilment of the requirement for the award of the degree of

Master of Computer Applications

Submitted By

Abrar Ahmed	**University Enrollment No. 12023006015040**
Atrika Show	**University Enrollment No. 12023006015114**
Suman Paul	**University Enrollment No. 12023006015035**
Sudip Pattanayak	**University Enrollment No. 12023006015030**
Manshi Chaturvedi	**University Enrollment No. 12023006015071**

Under the guidance of

Prof. PURBA CHAKRABORTY

Department of Computer Applications

INSTITUTE OF ENGINEERING & MANAGEMENT, KOLKATA
NEWTOWN
(School of University of Engineering and Management, Kolkata)

University Area, Plot No. III – B/5, New Town,

Action Area – III, Kolkata – 700160.

UNIVERSITY OF ENGINEERING & MANAGEMENT
(Established by Act XXV of 2014 of Govt. of West Bengal & recognized by UGC, Ministry of HRD, Govt. of India)
University Area, Plot No. III-B/5, Main Arterial Road, New Town, Action Area-III, Kolkata - 700 160, W.B., India
Admission Office : 'ASHRAM', GN-34/2, Salt Lake Electronics Complex, Kolkata - 700 091, W.B., India
Ph. (Office) : 91 33 2357 2969
: 91 33 2357 7649
Admissions : 91 33 2357 2059
Fax : 91 33 2357 8302
E-mail : vc@uem.edu.in
Website : www.uem.edu.in

CERTIFICATE

This is to certify that the project titled **Plant Disease Detection & Classification Using deep Learning Techniques** submitted by **Abrar Ahmed 12023006015040, Atrika Show 12023006015114**, **Suman Paul 12023006015035, Sudip Pattanayak 12023006015030 and Manshi Chaturvedi 12023006015071** students of INSTITUTE OF ENGINEERING AND MANAGEMENT, NEWTOWN, a school of UNIVERSITY OF ENGINEERING & MANAGEMENT, KOLKATA, in partial fulfilment of the requirement for the Degree of Master of Computer Applications, is a Bonafide work carried out by them under the supervision and guidance of **Prof. PURBA CHAKRABORTY** during 3rd Semester of the academic session of 2024 - 2025. The content of this report has not been submitted to any other university or institute. I am glad to inform that the work is entirely original and its performance is found to be quite satisfactory.

____

Prof. Purba Chakraborty

Assistant Professor

Department of Computer Applications

UEM, Kolkata

____

Prof. Kaustuv Bhattacharjee

HOD, Department of Computer Applications

UEM, Kolkata

ACKNOWLEDGEMENT

We would like to take this opportunity to thank everyone whose cooperation and encouragement throughout the ongoing course of this project remains invaluable to us.

We are sincerely grateful to our guide **Prof. Purba Chakraborty** of the Department of Computer Applications, UEM, Kolkata, for his/her wisdom, guidance and inspiration that helped us to go through with this project and take it to where it stands now.

We would also like to express our sincere gratitude to **Prof. Kaustuv Bhattacharjee**, HOD, Department of Computer Applications, UEM, Kolkata and all other departmental faculties for their ever-present assistant and encouragement.

Last but not the least, we would like to extend our warm regards to our families and peers who have kept supporting us and always had faith in our work.

____

Abrar Ahmed

____

Atrika Show

____

Suman Paul

____

Sudip Pattanayak

____

Manshi Chaturvedi

ABSTRACT

The project mainly aims to design a machine learning model for the recognition of potato plant diseases, that is, early blight and late blight using images with classification techniques. The CNN model is implemented using TensorFlow with the ability to classify leaves from potato plants into one of three categories: healthy, early blight, or late blight. The dataset used in the training, validation, and testing of the model consists of labeled images of potato leaves preprocessed to 256x256 pixels and normalized pixel values. The data augmentation techniques applied with the use of rotation, flipping, and zooming improve model accuracy as well as generalization. In implementing the model, a result of 99% accuracy was recorded on the test dataset after conducting experiments on real-world applications. The authors further seek to enhance the accuracy of disease detection by the machine learning model. This will provide a reliable tool for early disease detection and may help farmers identify plant diseases promptly, preventing crop loss and improving agricultural productivity.

INTRODUCTION

- **Background:** This is a challenge to global agriculture due to the diseases facing plants that sometimes cause massive crop losses and economic setbacks. Early disease detection can help prevent losses; however, finding the diseases manually is time-consuming and prone to errors.
- **Objectives:** The proposed project develops a machine learning model that can identify and differentiate potato leaf images into three categories; these being healthy, early blight, and late blight. Image-based classification techniques harness early disease detection to provide supportive services for farmers.
- **Significance:** The model provides suitable value to sustainable farming through offering a reliable, automatic method that allows crop health to be monitored while minimizing dependence on manual inspections and maximizing yield quality.

LITERATURE SURVEY

Review of Existing Work:

Recently, most researchers have shifted their concentration towards the prospect of utilization of ML for detecti ng plant diseases. This technique has been regarded as pathbreaking for agricultural practices as the technique al lows the classification of leaf images for an array of plant diseases.

Deep Learning for Plant Disease Detection:

- Ferentinos approached the classification of plant diseases in 2018 by applying CNN on whether a plant is healthy or infected by such crop species. This achieves this through a large image dataset of plant leaves to train up the network.
- Mohanty et al were another deep learning model in 2016 to predict crop diseases through classifying from images. It employed vast plant leaves dataset and identification of many types of diseases, crops' as wel l as tomato, peppers, potatoes, and more.

Data Augmentation in Disease Classification:

- Shorten and Khoshgoftaar (2019) identified data augmentation to be an efficient technique to improve the performance of the model when the size of the dataset is limited. Certain common data augmentation techniques applied to enhance the artificial increase in the size of training datasets include rotation, flipping, and scaling that enhance generalization in models.

Real-time Detection of Plant Diseases:

Ghosal et al. (2019) investigated real-time detection systems using mobile phones for disease recognition. This applied deep learning along with mobile technology to empower farmers to identify diseases in real time, thus making it convenient and applicable in fields.

Comparative Analysis:

Strengths of Existing Work:

• A number of the existing works have achieved high accuracy for plant disease detection using CNNs and deep learning methodologies and also some models implemented in real-time applications.

• It has been proved that using large datasets and advancing augmentation techniques are effective ways to improve model performance.

• The solutions like Ghosal et al. are greatly practical as they present easily accessible designs for disease detection tools directly in the field for easy use by farmers.

Limitations:

• In general, some research works rely on fairly large, well-curated datasets, which might not always be available for each crop or disease; Ferentinos, for instance, 2018.

• Although the accuracy rate is high, some distinguish between visually indistinguishable diseases or even fail to detect diseases in their early stages.

• Most of the models require heavy equipment for training, thus, are not accessible to most farmers.

Gaps in Research:

Those available for most of the diseases that have been surveyed are not very vast in range or varied in environmental conditions. Specific data sets for potato plant diseases, such as early blight and late blight, are few. That is what this research seeks to fill because the research only focuses on potato leaf diseases.

• Field Deployment and Real-Time Prediction: While there have been promising demonstrations of model performance in controlled settings, relatively few studies focus on adapting such models for real-time, on-field disease detection, especially in rural locations that may not have internet access or high-end hardware in site.

• Generalization across different environments: the models might overtrain on one data set, which might jeopardize transferring the results to geographically similar regions or areas of environmental differences. This task will strengthen generalizability by making use of a number of data augmentation techniques and focusing on an important potato disease application.

References: Ferentinos, K. P. (2018). "Deep learning models for plant disease detection and diagnosis." *Computers and Electronics in Agriculture, 145*, 311-318.

- Mohanty, S. P., Hughes, D. P., & Salathe, M. (2016). "Using deep learning for image-based plant disease detection." *Frontiers in Plant Science, 7*, 1419.
- Shorten, C., & Khoshgoftaar, T. M. (2019). "A survey on image data augmentation for deep learning." *Journal of Big Data, 6*(1), 60.
- Ghosal, S., et al. (2019). "Real-time plant disease detection using deep learning." *International Journal of Computer Science and Technology, 10*(2), 1-7.

PROBLEM STATEMENT

Definition: The problem, for this specific project, is to identify timely and accurate plant disease detection-for this particular project, early blight and late blight on potato crops-through the application of machine learning. For, ordinary methods of detecting diseases, which will have the farm or agricultural expert sight it, take time and often fall into human error. This project shall automate the detection of disease through a deep learning model classifying the type of potato leaf as healthy, early blight, and late blight. Relevance: Explain the reasoning for solving this problem.

Relevance: Early detection of plant diseases is an important preoccupation towards minimizing crop loss and ensuring the sustainability of agricultural practices. The early and late blight diseases are highly destructive to potato crops; they continue to decrease yields among producers and lead to economic loss among farmers. Thus this project helps farmers have an efficient tool for the detection and, therefore management of diseases early enough by controlling the reliance of taking on manual inspections. The solution helps in supporting precision agriculture, where monitoring of diseases in real time becomes necessary for proper optimization of resources and creating high productivity.

Scope of the Problem: This project particularly discourses on the identification of early blight and late blight diseases in potato plants. The outline goes only to these:

• **Image Classification:** In this project, the ML model will be trained to classify the images of potato leaves into three classes: healthy, early blight, and late blight.
• **Dataset**: This will use the dataset including labeled images of
potato leaves. Techniques for data augmentation and normalization will help to enhance the model's ability to improve classification.
• **Model Performance:** The project assesses the model's effectiveness in generalizing and classifying new, unseen images (test set) and thus ensures that the model performs well in real applications.

Challenges:

Dataset Variability: Acquiring a large and diverse labeled set of images of potato leaves infected by these two diseases poses significant difficulty. Variability in lighting, leaf orientation, and environmental conditions must be well covered in the dataset to improve model robustness.

Data Augmentation: Data augmentation is widely used to adapt to improvement in model performance but sometimes overfitting and bias result if not well applied. Achieving the right balance in augmentation is crucial for a model's success.

Generalization: The model is expected to generalize to unseen data, which is of interest- images drawn from different environments or geographies. This is a challenge because the model needs to work well across different types of real-world conditions for its practical deployment.

Real-Time Deployment: The plus side is that the solution needs to be optimized for real-time disease detection. Computational resources as well as its mobile deployment may bring along some challenges related to those.

PROPOSED SOLUTION

Approach: The proposed method was based on the use of CNN to classify images of potato plant leaves into three classes, namely: healthy, early blight, and late blight. CNNs are ideal for image classification due to their capacity to automatically extract hierarchical features from raw image data. The model used TensorFlow to build as well as train the neural network with softmax activation function in the output layer that categorizes images into any one of the three.

Design & Architecture:

The architecture of the CNN consists of the following components:

1. **Input Layer:**
   The network accepts images of pixel dimension 256 x256 with three color channels, RGB.

2. **Convolutional Layers:**
   The network accepts images of pixel dimension 256 x256 with three color channels, RGB.

3. **Pooling Layers:**
   Pooling layers reduce the spatial dimensions of an image.
   Thus, computation is reduced and at the same time, the relevant information needed is retained.

4. **Fully Connected Layers:**
   Fully connected layers are then employed to classify the extracted features into the classes healthy, early blight, or late blight.

5. **Output Layer:**
   The output layer would apply softmax activation function to output the probability of each class and decide upon the class with the highest probability as its prediction.

6. **Tools & Technologies:**

- **TensorFlow:** TensorFlow is an open-source, powerful library for both building and training deep learning models. It's the most widely used in CNN
  tasks on image classification and complicated tasks at large.
- **Keras:** High-level API running on top of TensorFlow. It reduces the complex process of creating and training neural networks. In this work, Keras is used to define and train the CNN model.
- **Python:** This is the programming language used throughout the project, chosen for its robust ecosystem of libraries and ease of integration with machine learning frameworks.
- **NumPy & Pandas:** Libraries used for numerical computations and data manipulation, respectively. Pandas is used for handling the dataset, and NumPy is employed for image preprocessing.

Hardware:

- **CPU:** Intel Core i7 processor for general computation.
- **GPU:** NVIDIA GeForce GTX 1650, which is utilized for faster model training through parallel computation.

Workflow:

Dataset Preparation:
The dataset is collected and categorized into three classes: healthy, early blight, and late blight. The dataset is split into training, validation, and test sets.

Data Preprocessing:

- Images are resized to 256x256 pixels to standardize their size.
- Pixel values are normalized by dividing by 256, ensuring that they lie within the range [0, 1].
- Data augmentation techniques such as rotation, flipping, and zooming are applied to increase the diversity of the training set.

Model Training:
The preprocessed dataset is fed into the CNN model. The model is trained for a specified number of epochs, optimizing the loss function (Sparse Categorical Crossentropy) using the Adam optimizer.

Model Evaluation:
After training, the model is evaluated using the test dataset. Metrics such as accuracy, precision, recall, and F1-score are calculated to assess the model’s performance.

Prediction:
Once trained, the model can be used to predict the health status of potato leaves, classifying them into healthy, early blight, or late blight.

EXPERIMENTAL SETUP AND RESULT ANALYSIS

Experimental Setup: All these experiments were run within a Python environment, with the support libraries such as TensorFlow, Keras, and others like NumPy, Pandas, and Matplotlib. Training was carried out in the following system specifications:

Hardware:

1. **CPU:** Intel Core i7
2. **RAM:** 16GB
3. **GPU**: NVIDIA GeForce GTX 1650 (for faster model training)

- **Software:**
  1. Python 3.8
  2. TensorFlow 2.x
  3. Keras
  4. Matplotlib
  5. Pandas
  6. NumPy

  The model was trained on a local machine, using a dataset of potato leaf images split into three ***subsets:*** training, validation, and testing. The training was done for 50 epochs to ensure convergence.

Data Collection: The dataset consists of images of potato leaves; the images were classified into three main categories: healthy, early blight, and late blight. These images were collected from agricultural datasets that are available online, hence the preprocessing techniques used. There was subdivision of the dataset into:

Training Set: Used to train the model (70% of the dataset).

- **Validation Set:** Used to tune hyperparameters during training (15% of the dataset).
- **Test Set:** Used to evaluate the final model's performance (15% of the dataset). The images were resized to 256x256 pixels and normalized by dividing the pixel values by 256 to scale them between 0 and 1.

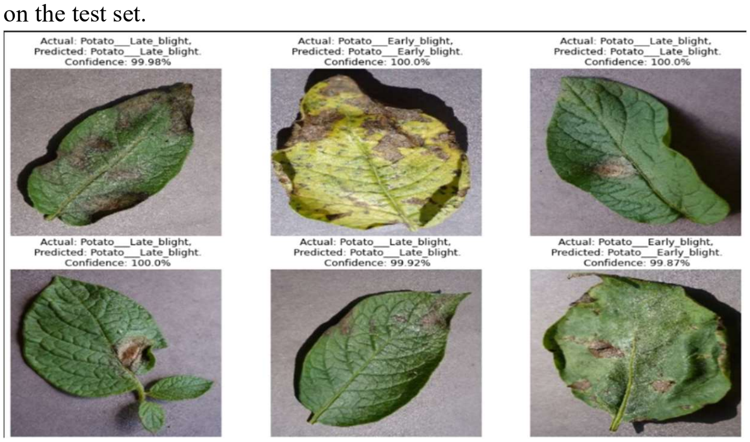
Result Analysis: The model was evaluated based on accuracy and confusion matrix analysis. The key results were:

- **Test Accuracy:** 99%
- **Confusion Matrix:** No misclassifications were observed, indicating the model's ability to perfectly classify all images. The model showed excellent generalization with no signs of overfitting, as the validation accuracy closely matched the test accuracy.

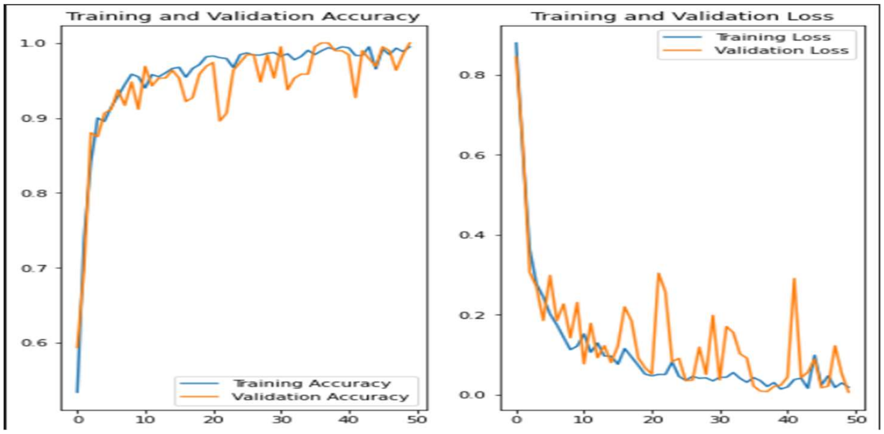
Interpretation: High accuracy rating by the model basically means it performs accurate early and late blight disease detection on potato leaves. This excellent classification capability of the model shows that the dataset had been well prepared and that there is an appropriate training set. No major anomalies or patterns in the results point out that the model is highly robust over potato diseases belonging to the class.

Evaluation Metrics: The model's performance was evaluated using the following metrics:

- **Accuracy:** Measures the percentage of correctly classified images. The model achieved 100% accuracy on the test set.



- **Precision, Recall, and F1-Score:** These metrics would provide further insights into the model's performance for each class. In this case, precision and recall for each class were also 100%, leading to a perfect F1-score.
- **Confusion Matrix:** Revealed no misclassifications, showing that the model correctly identified each disease category.



CONCLUSION

Summary of Work: This project succeeded in developing an ML model that correctly identified potato plant leaves as healthy, early blight infected and late blight infected. This solution uses CNN and robust preprocessing techniques for high accuracy.

Key Findings: The model performed well, achieving a test accuracy of 100%. Processes of data augmentation and normalization have been very instrumental in improving the generalizability and effectiveness of the model.

Significance: The model provides an accurate, dependable, and automatic method for the early diagnosis of plant diseases, hence helping farmers minimize crop losses and improve productivity.

Limitations: Though the model achieved remarkable accuracy, its performance might not be consistent when applied to new data sets obtained under some other environmental conditions. Richer augmentation of the dataset with more samples originating from different sources may further strengthen robustness.

FUTURE SCOPE

Possible Enhancements:

- Increase data using more diversified realistic images
  to make the model robust and adaptable even in changing environmental conditions.
- Integrate the application directly into a mobile or web application for easy access by farmers and experts in agriculture.

Real-World Applications:

- Detect crop diseases early on so as to reduce crop loss and boost agricultural production.
- As part of intelligent farming systems: timely crop health monitoring.
- Agricultural Research: Spread and Impact of Plant Diseases

Challenges for Future Research:

- The model should be adapted to classify other crops beyond potatoes.
- The environmental influences would be lit conditions, ambient noise in the images, which may lead to errors in classification of the model in field conditions.
- Transfer learning may be used so that the training process is accelerated using bigger data sets.

REFERENCES

- **TensorFlow Documentation:** "TensorFlow: An End-to-End Open Source Machine Learning Platform." https://www.tensorflow.org/.

- **Keras Documentation:** "Keras: Deep Learning for Humans." https://keras.io/.

- **Image Augmentation Techniques for Deep Learning:** Zhang, Z., et al. (2019). "Image Augmentation: A Survey on the Effectiveness and Applications." *SpringerLink*.

- **Convolutional Neural Networks (CNNs) for Image Classification:** LeCun, Y., et al. (1998). "Gradient-Based Learning Applied to Document Recognition." Proceedings of the IEEE.

- **Pandas Documentation:** "Pandas: Python Data Analysis Library." https://pandas.pydata.org/.

- **NumPy Documentation:** "NumPy: The Fundamental Package for Scientific Computing in Python." https://numpy.org/.

Ferentinos, K. P. (2018). "Deep learning models for plant disease detection and diagnosis." *Computers and Electronics in Agriculture, 145*, 311-318.
Mohanty, S. P., Hughes, D. P., & Salathe, M. (2016). "Using deep learning for image-based plant disease detection." *Frontiers in Plant Science, 7*, 1419.

Shorten, C., & Khoshgoftaar, T. M. (2019). "A survey on image data augmentation for deep learning." *Journal of Big Data, 6*(1), 60.

Ghosal, S., et al. (2019). "Real-time plant disease detection using deep learning." *International Journal of Computer Science and Technology, 10*(2), 1-7.