Potato Plant Disease Detection System using Deep Learning

Dr. Shambhu Rai
Project Guide
Bharati Vidyapeeth's Institute of Management and Information Technology
Navi Mumbai, India
shambhumca1@gmail.com

Tushar Vagh
Student
Bharati Vidyapeeth's Institute of Management and Information Technology
Navi Mumbai, India
trvagh123@gmail.com

***Abstract*—Agriculture serves as the cornerstone of global food security, sustaining billions of people across the world.and driving economic growth. However, plant diseases may negatively affect crop production, leading to financial difficulties for farmers. For these reasons, Consequently, crop yields are severely compromised, pushing farmers into economic hardship. Therefore, identifying plant diseases at an early stage is critical for effective crop management and enhancing overall agricultural output.**

In recent years, rapid progress in artificial intelligence and deep learning technologies has paved the way for the creation of automated image-based systems capable of recognizing plant diseases. The study introduces an AI-powered web-based application AgriDiagnose that utilizes CNN to identify disease in potato plants. The model's results are then returned to the user through a simple web interface, along with recommended treatments. The method demonstrates how deep learning can benefit agriculture through improved early detection of disease and assisting farmers in making better decisions regarding their crops.

Farmers now have access to more digital technologies than ever before. This has led to the creation of automated smart systems that help farmers keep track of their crops and provide help with diagnosing problems with those crops. Research is currently being conducted into the use of pictures taken of plant leaves using modern smartphones, which allow farmers to capture high-quality images of their crops directly within their fields.

Agricultural support systems can provide both real-time data analysis and the ability to suggest optimum choices to users by applying deep learning models together with web-enabled applications. By using these systems, manual verification steps for crop disease diagnosis are reduced, and the rate at which accurate diagnoses are made increases.

I. INTRODUCTION

Agriculture is crucial for the survival of the world's population and for continued economic growth, especially in developing countries where a significant portion of the population relies on agriculture for their livelihoods [4]. Plant diseases are one of the primary components that affect agricultural productivity. Plant diseases have the ability to spread rapidly across fields, which often leads to material loss associated with lower quality produce, reduced yields and ultimately economic hardship for farmers [1].

Of the various crops produced across the globe, potatoes are ranked among the top three in total production by volume and have high susceptibility to a number of different fungal and bacterial infectious diseases. These two diseases can severely damage the above-ground portions of the potato plant by altering leaf anatomy and decreasing the ability of the leaf to photosynthesize as well as increasing yield loss [7]. Manually inspecting potato plants or consulting with an agricultural expert are the two most common ways for potato farmers to identify potential plant diseases.

Manual inspection of the plants takes time, is often subjective and provides inconsistent accuracy especially when the symptoms for one disease might resemble another disease's symptoms [8].

The proposed system, AgriDiagnose, is an AI based web-based application that allows the user to identify whether a potato leaf is healthy or infected with a specific disease using deep learning techniques. The user will submit images of potato leaves for the CNN model to evaluate the health status of the potatoes.

A. Motivation

- Plant diseases typically go unidentified until too late, resulting in loss of the crop.
- Farmers use conventional methods to identify their plant diseases.
- Rural areas typically do not have access to the expertise of agricultural professionals.
- Certain plant diseases may exhibit similar visual characteristics.
- The resources for farmers to utilize when identifying diseases are often not accessible to those who are small-scale farmers.
- Digital solutions that use a simple image of the leaf on the plant are needed for use in identifying disease.
- AI technology has demonstrated a potential for success in the identification of numerous plant diseases.
- Systems for the real time detection of disease could assist growers with controlling plant diseases and thereby increasing productivity.

B. Objectives

- Establish the use of an intelligent system for identification of potato diseases.
- Develop a method for preprocessing of image data.

- Incorporate the model into a web application based upon the Flask framework.
- Generate recommendations for the control of identified diseases.
- Develop a simplistic user interface.
- Reduce the time required for determining plant disease.
- Improve agricultural productivity.

II. REVIEW OF LITERATURE

The literature on plant disease identification and crop surveillance has been examined by numerous researchers who applied artificial intelligence and machine learning techniques toward their goals. The following are three areas of required research, along with the associated references:

A. Deep Learning Approaches

Plants have unique characteristics associated with their different growth patterns such as the number of branches, leaves, and roots. Deep learning methods like Convolutional Neural Networks automatically learn various characteristics related to plant appearance such as edges, textures, and disease patterns, and perform very well in identifying diseases within a particular plant [1], [2], [9]–[11].

B. Traditional Machine Learning

The advancement of artificial intelligence and machine learning has led to the emergence of novel approaches that rely on manually engineered features — including color distributions, texture patterns, and shape representations — extracted from previously documented diseases, and processed through conventional machine learning techniques such as Support Vector Machines and Decision Trees [7], [8].

C. AI-Based Agricultural Systems

A growing number of scholars have started integrating image processing methodologies with deep learning architectures to engineer more precise and reliable systems for the identification of plant diseases [**?**], [14]. Researchers have used techniques such as edge detection, segmentation, and color analysis to create more suitable im-ages for use in conjunction with machine learning algorithms [8]. Numerous recent studies emphasize the critical importance of having large, diverse datasets to create strong deep learning networks and the availability of publicly accessible plant image databases has accelerated progress regarding the de-velopment of plant disease identification technologies [3], [13].

Even though these advancements are happening, there are still many models still stuck in the lab. They have not yet been able to transition from lab to on-farm use. More and more, agronomists are looking for places to apply these trained models in ways that enable user-friendly interactions while providing real-time predictions of disease occurrence [4].

III. PROBLEM STATEMENT

Plant disease is a major contributor to reduced yields and food security in agriculture. Farmers are often having difficulty determining what problem they have due to a lack of access to expert knowledge or diagnostic tools. Disease detection is often delayed, leading to rapid outbreaks in fields. Manual inspection is typically slow and does not necessarily lead to an accurate diagnosis.

Therefore, an automated and simplified detection method is needed to allow farmers to accurately identify plant disease through the use of images taken by their mobile phones.

Another major issue is that many farmers do not know about modern digital solutions to help manage crops. When technologies are available, they may not be used due to access or complexity of the system.

IV. DEVELOPED FRAMEWORK FOR AUTOMATED POTATO PLANT DISEASE IDENTIFICATION

A. System Architecture

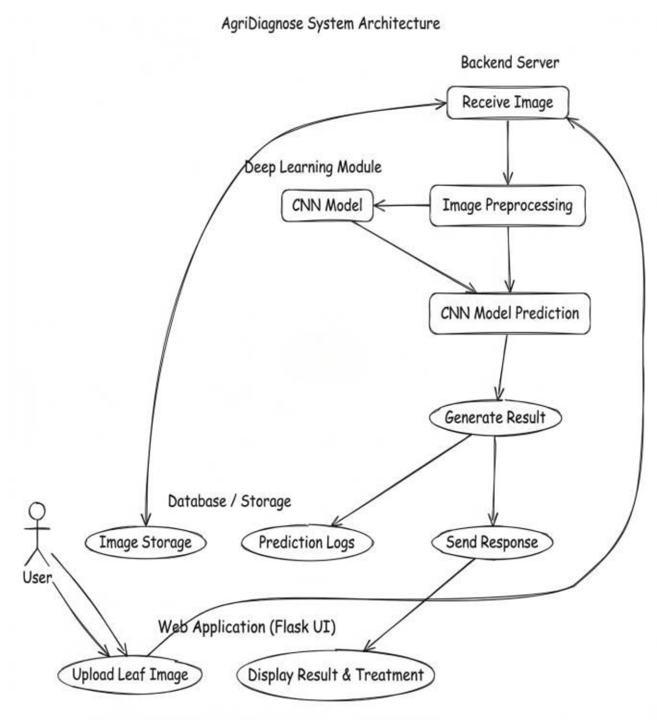

Fig. 1. AgriDiagnose System Architecture showing data flow from user input to CNN-based prediction and result display through the web application.

B. Functional Modules

- Image Preprocessing Module
- CNN Disease Detection Module
- Model Training and Evaluation Module
- Treatment Recommendation Module
- Flask Web Application Module

V. METHODOLOGY

To collect images of potato leaves will be done, whose health status will be tested for training of the deep learning model. The leaf images will be prepared using resizing, normalization and augmentation methods. The CNN model will learn to recognize visual patterns that correlate to plant disease. The user will use a leaf image through the application interface to obtain disease class predictions. The resulting output information will be displayed with treatment suggestions.

VI. DATASET DESCRIPTION

A. Image Dataset

TABLE I
FIELD DATA COLLECTION FOR POTATO CROP

Location	Experience	Observed Symptoms	Irrigation
Dhule	12+ yrs	Yellowing leaves, brown spots, rapid spread in humidity conditions	Borewell + Drip
Nashik	15 yrs	Black patches on leaves, leaf curling, disease starts from lower leaves	Flood Irrigation
Jalgaon	10 yrs	Dry spots, wilting during sunlight, pest damage observed	Drip Irrigation
Pune	9 yrs	Yellow edges on leaves, gradual drying and weakening of plant	Borewell
Satara	11 yrs	Brown circular lesions, spreading spots on leaf surface	Drip Irrigation
Ahmednagar	13 yrs	Leaf curling with black fungal patches and discoloration	Canal Irrigation
Kolhapur	8 yrs	Wilting leaves, holes caused by pests and insects	Rain-fed
Solapur	14 yrs	Yellowing of leaves, dry patches, reduced growth rate	Borewell
Latur	7 yrs	Brown spots, leaf damage, early signs of infection	Drip Irrigation
Aurangabad	16 yrs	Black patches, fast disease spread during humid conditions	Canal + Borewell

B. Dataset Split

Training Set: 70% Testing Set: 30%

The images contained in the dataset are taken in multiple environments and contain a range of conditions such as different lighting circumstances and different backgrounds that provide diverse content. All this variety helps the model to learn many strong visual features and thus perform well on new images when predicting their labels [3], [13].

Collecting such a variety is essential to build the right model to classify images correctly because we must also include images representing each disease at all stages of disease development to ensure they can be classified when visual recognition is typically more difficult.

The images are organized into separate folders of each disease's images, providing a structured way for the model to learn how to classify images and making it much more efficient to load data from those respective folders into memory for rapid learning.

VII. DATA PREPROCESSING

A. Image Cleaning

- Removal of unwarranted images
- Verification of appropriate class labels for each image

B. Image Processing

- Resize original images to dimensions of 224×224 px
- Normalize pixels in original images

C. Data Augmentation

- Rotation
- Horizontal flip
- Zoom

The preparation of visual data prior to model training constitutes a fundamental step in elevating the overall performance of machine learning systems. As an illustration, raw captured images frequently contain unwanted noise, the illu-mination in the original image will vary greatly depending on the source, and each image will have background information that is not relevant to the machine learning process. All of these problems will need to be eliminated with image cleaning and pre-processing techniques for the model to be accurate in labelling images.

Normalizing the pixel values of an image to stay within a consistent set of numbers allows smoother learning for the neural networks. Resizing images to uniform dimensions also helps the CNN model to accept all images that go into it.

Data augmentation techniques allow an increase in diversity of image datasets by creating alternative varieties of the same image dataset. This reduces overfitting and allows the model to perform better when making predictions on previously unseen data.

VIII. DEEP LEARNING MODEL

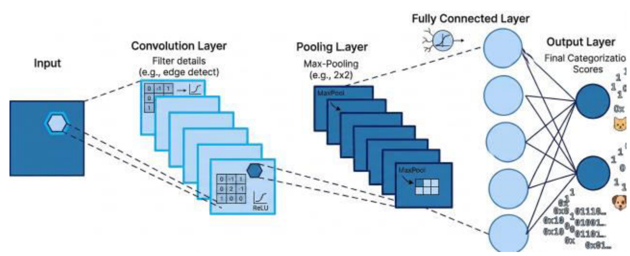

Fig. 2. CNN Architecture for Plant Disease Detection

The architecture of the Convolutional Neural Network (CNN) is composed of the following layers: convolutional, activation, pooling and fully connected layers which make up the complete classification component of the model [5], [6], [12].

IX. TRAINING STRATEGY

Training configuration consists of:

- Loss Function: Cross-Entropy
- Optimizer: Adam
- Train–Test Validation

Throughout the training phase, the convolutional neural network progressively acquires an understanding of the underlying associations between the visual representations of plant diseases and their corresponding internal parameters using a technique called backpropagation [5], [12]. An optimizer algorithm iteratively modifies the weights of the model to achieve an optimal value for the loss function during this process.

By utilizing many epochs for training, the learning of progressively increased accuracy will ultimately produce a high level of correct classification from the model The developed model will be evaluated against an independent dataset, distinct from the one utilized during training, with the dual purpose of assessing its overall predictive performance and mitigating the risk of overfitting to the training data.

Hyperparameter values like learning rate, number of epochs, or batch size have a significant effect on improving or decreasing how effective the training procedure is.

X. SYSTEM WORKFLOW

1) User uploads the image of the potato leaf.
2) A series of systematic image preprocessing operations are applied to the submitted visual input.
3) The CNN classifier analyses the image.
4) The classification of the disease detected by the CNN model is generated.
5) The system displays the user a classification of the image submitted and a list of treatments based on the classification made by the CNN model.

The AgriDiagnose system has been designed to have a simple user interface. Farmers, in the case of an image of their potato plant, or general users can submit an image via a web interface without having knowledge about programming or machine learning.

Once the image is submitted via the web interface, the backend server will take care of pre-processing the image and transfer the pre-processed image to the trained CNN model. Once the image is processed by the trained CNN model, it registers a prediction for the disease class detected from the image submitted by the user.

Value is given to the user for the classification of the image by searching the database for treatment for that disease and displaying the treatment with the prediction generated. This way it is easy for users to see how their plant is, and get the required information to take action based on the health of their plant.

XI. EVALUATION METRICS

Among the various metrics employed to assess the effectiveness of classification models, accuracy remains the most widely recognized and frequently utilized measure [2], [11]. However, accuracy does not always tell the complete story about a model's performance; this is especially true for evaluating performance on an imbal-anced dataset.

Precision measures how many correctly predicted examples of a positive value were found from all predicted examples of a positive value. Recall is a measure of how well a model identifies actual positive values within the sample.

XII. Concluding Remarks and Prospective Research Directions

The use of AI for automating the process of detecting plant diseases, using deep learning methods through the AgriDiagnose System has potential to aid in the early identification of plant diseases by providing accurate diagnosis using pictures of leaves, and then supporting the production of treatment recommendations [1], [2], [14].

In future implementations there will be an increase in the dataset size that will include more crops, as well as the development of a mobile application to expand the access of the system.

The integration of Artificial Intelligence within the agricultural sector signifies a significant step towards modernizing farming practices, with respect to how crops are monitored and the need for growers to perform visual inspections of their own crops.

In the future, the AgriDiagnose system can be expanded by integrating additional crop diseases and incorporating real-time monitoring features. Mobile applications and cloud deployment can further improve accessibility and scalability of the system.

Advancements in artificial intelligence, combined with the increasing availability of agricultural datasets, will continue to drive innovation in smart farming technologies [4]. These technologies have the potential to improve global food security and support sustainable agricultural practices.

References

[1] S. P. Mohanty, D. P. Hughes, and M. Salathe´, "Using deep learning for image-based plant disease detection," *Frontiers in Plant Science*, vol. 7, pp. 1–10, 2016.

[2] K. P. Ferentinos, "Deep learning models for plant disease detection and diagnosis," *Computers and Electronics in Agriculture*, vol. 145, pp. 311–318, 2018.

[3] J. G. A. Barbedo, "Impact of dataset size and variety on plant disease classification using deep learning," *Computers and Electronics in Agriculture*, vol. 153, pp. 46–53, 2018.

[4] A. Kamilaris and F. X. Prenafeta-Boldu´, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, 2018.

[5] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016.

[6] A. Krizhevsky, I. Sutskever, and G. Hinton, "ImageNet classification with deep convolutional neural networks," *NIPS*, pp. 1097–1105, 2012.

[7] R. P. P. Singh and M. Misra, "Plant leaf disease detection using image processing techniques," *IJSETR*, vol. 8, no. 4, 2017.

[8] P. Revathi and M. Hemalatha, "Classification of cotton leaf spot diseases using image processing," *IEEE ICETEST*, 2012.

[9] S. Sladojevic et al., "Deep neural networks based recognition of plant diseases by leaf image classification," *Computational Intelligence and Neuroscience*, 2016.

[10] H. Brahimi, A. Boukhalfa, and A. Moussaoui, "Deep learning for tomato disease classification," *IEEE*, 2017.

[11] A. Too, L. Yujian, S. Njuki, and Y. Yingchun, "A comparative study of fine-tuning deep learning models," *Computers and Electronics in Agriculture*, 2019.

[12] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436–444, 2015.

[13] D. Hughes and M. Salathe´, "An open access repository of plant images," *arXiv*, 2015.

[14] S. Picon et al., "Deep learning for plant disease detection and classification," *Biosystems Engineering*, 2019.