A Plant Disease Detection App Using Machine Learning

Sarthak Kumar Singh[1], Shivam Bhardwaj[2], Dr. Riman Mandal[3]
[1,2]*Department. of Computer Science and Engineering, Galgotias University, Greater Noida, India*
[3]*Professor, Galgotias University, Greater Noida, India*

Abstract—**Plant diseases cause great reduction in agricultural production which in turn affects food security and farmers' income. It is of great importance to detect them early and with precision to reduce crop damage and at the same time improve product quality. We present a machine learning based solution for the detection of plant diseases which we have implemented via Convolutional Neural Networks (CNN). The put forth system does image acquisition, preprocessing, feature extraction and classification which in turn leads to accurate identification of plant diseases. We used the Plant Village data set for our experiments which we report to perform better than do traditional machine learning methods. Also, we have developed the system as a web-based application which enables real time disease diagnosis in the field.**

Index Terms—**Plant Disease Detection, Machine Learning, Deep Learning, Convolutional Neural Networks, Agriculture, Image Processing**

I. INTRODUCTION

Most developing countries' economies are heavily reliant on the agriculture sector, which also supports the food security needs of the world's growing population. Agriculture as a sector faces many challenges, including the prevalence of plant diseases that not only reduce the volume of food crops but also threaten the quality and productivity of the food. Agricultural scientists respond that the principal cause of a 30-40 percent loss of crop yields is from the lack of adequate and timely control of crop diseases. Therefore, for agriculture to be sustainable, the control of plant diseases is essential.

Traditionally farmers and human experts diagnose plant disease by visually inspecting crop samples. The manually visual assessment process is slow, prone to error due to human uncertainties, requires excessive labor and time. If farmers are distant from relevant advice and know-hows, delayed diagnosis may result in applying ineffective treatment measures thus increasing the production costs, damaging the environment and risking crop loss.

In recent past we have seen great advances in the fields of deep learning, machine learning and vision as they apply to agriculture and image diagnosis which in turn has brought about new solutions for the automatic recognition of plant diseases from their visual symptoms with minimal human input. Visual symptoms on the plant leaves such as that of discoloration, spots, textural and hole effects may be very accurately and quickly assessed by machine learning models. Deep learning is at the fore front of machine learning methods used in large scale feature extraction. What we see in deep learning models in particular Convolutional Neural Networks is that they are able to learn to represent complex feature hierarchies in raw images out of which any need for intermediate image preprocessing or in toto feature selection is done away with.

This project we are developing a plant disease identification system which uses deep learning to classify plant leaf images. The students' goal is to present a reliable and user-friendly web-based application which helps farmers and related professionals in early-stage disease management.

II. RELATED WORK

A significant amount of research has explored automated plant disease diagnosis via image processing and machine learning. Early in the field of study we looked at what is today considered traditional image analysis methods which included the use of certain features from leaf images like color histograms, texture descriptions, and geometric shape properties. We then put these features into classifiers which at the time included Support Vector Machines (SVM), k-Nearest Neighbours (k-NN), Naive Bayes, and Decision Trees. Also, in that which. methods showed good accuracy in controlled experiments, their performance dropped in real-world settings, where lighting, backgrounds, and imaging factors can vary greatly.

With the rise of deep learning we have seen a trend in recent research which is a shift towards the use of CNN based models for the task of identifying plant diseases. CNN's are able to extract complex features from raw images which in turn makes them very effective for visual recognition. Also, Mohanty et al. applied deep CNN models to the Plant Village database and reported very strong results which were consistent across many different crops. Also it was shown by Ferentinos that deep learning platforms are indeed very reliable and scalable for plant disease detection. Also a number of studies looked at transfer learning which is a practice of taking pre trained models like VGG16, ResNet, and Inception and fine tuning them for plant disease datasets. This is a very useful approach which reduces training time and at the same time increases accuracy in particular when you have limited labeled data. At the same time though these transfer learning techniques require large amounts of computing power. which can limit their use in areas with fewer resources.

Overall in research CNN based methods are seen to outperform conventional machine learning tools in terms of accuracy and reliability. That said we still have a great need for simple to use systems which at the same time provide effective disease and practical solutions. This study we set out to address that need by developing a CNN based plant disease detection app for use in real world agricultural settings. often these require large amounts of computing power.

In also which other studies have looked at plant disease diagnosis via classical image processing and machine learning. In the early days we saw research which reported on the use of manual feature extraction which included color, texture and shape descriptors also in which they used traditional classifiers such as SVM, k-NN, and Decision Trees. While they did report moderate results these methods' great dependence on specialized feature engineering was a issue as well as their very environmental sensitivity.

The advancements made in deep learning have improved the effectiveness of deep learning in classification tasks. Mohanty et al. showcased the successes of deep CNN models for plant disease classification on the Plant Village dataset. Similarly, Ferentinos validated the effectiveness of deep learning models across various crops. Collectively, these works underline the potential of the CNN-based approach for large-scale, reliable, and robust plant disease diagnosis.

III. SYSTEM ARCHITECTURE

The proposed plant disease detection system uses a modular, layered structure that allows for scalability, reliability, and effective image data processing. The structure includes several main modules: image acquisition, image preprocessing, CNNdriven feature extraction, disease classification, and result visualization.

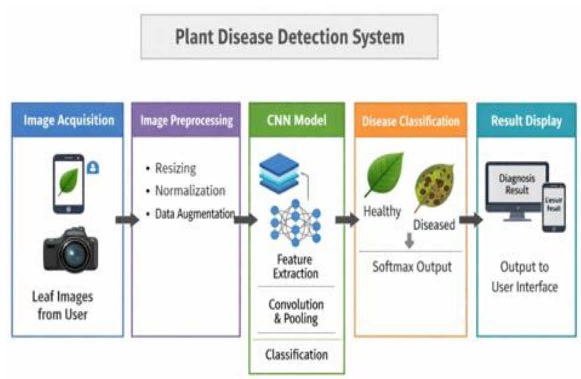

Fig. 1. Overall system architecture of the plant disease detection application

In at the image acquisition stage we have that users use mobile devices and digital cameras to take in plant leaf images.

As for the preprocessing stage we do resizing, normalization, and data augmentation in order to get the images ready for analysis. We use a CNN model which puts out the Softmax which in turn does the classification of the features. At last the results of the diagnosis are put out via a web or mobile based interface.

IV. PROPOSED METHODOLOGY

The proposed method describes the full workflow for automatically detecting plant diseases with machine learning. The system processes leaf images, extracts important features, and accurately identifies different plant diseases.

A. Dataset

The Plant Village dataset we use for training and evaluation has a collection of thousands of labeled images of healthy and diseased leaves from many crop species. This which also has great diversity and a good balance of classes in which makes it a great fit for supervised deep learning models.

B. Image Preprocessing

Before training the model, we preprocess the images to improve data uniformity and quality. We resize all images to 224×224 pixels and normalize them so that pixel intensities are between 0 and 1. We apply data augmentation operations, such as rotation, flipping, and zooming, to improve robustness and reduce the risk of overfitting.

C. CNN Model Architecture

The CNN architecture we use has several convolutional layers which apply ReLU activation functions to identify out of the spatial patterns in the leaf images. Also we have max pooling layers that reduce the dimension of feature maps which at the same time preserve what is important. Also we see the use of drop out layers which in turn reduces over fitting. At the end we have full connected layers that do the disease classification via a Softmax activation function.

D. Training Process

The network employs the Adam optimization algorithm and the loss function is categorical-cross entropy. The training is carried out in multiple epochs. Tracking performance, supporting convergence and promoting generalization is done with the help of validation data.

V. ALGORITHM DESCRIPTION

This section outlines the procedural steps used for detecting plant diseases.

A. Algorithm 1: Plant Disease Detection

[H] Plant Disease Identification Based on CNN [1] Acquire a leaf image from the user Perform image resizing and normalization Apply data augmentation techniques Use the CNN to extract image features Identify the disease class using a Softmax output layer Present the predicted result to the user

VI. TIME COMPLEXITY ANALYSIS

The time complexity analysis assesses the efficiency of the proposed system for detecting plant diseases. The system mainly relies on Convolutional Neural Network (CNN) for feature extraction and classification which is also the most computationally intensive part of the system.

The test results indicate that the CNN-based method outperforms the already proven classical machine learning techniques because it learns the significant characteristics by itself. The effectiveness of deep learning in the diagnosis of plant diseases is exhibited in the high level of accuracy obtained. The usability and accessibility for the farmers is also enhanced by the web-based application

A. Image Preprocessing Complexity

Let n be the number of input images and p be the number of pixels of each image. Preprocessing procedures include resizing, normalization, and augmentation, which are performed on each individual image. As a result, the time complexity for the preprocessing stage becomes:

$O(n\times p)$

B. CNN Feature Extraction Complexity

A CNN has several layers of convolutions. Let k denote the number of convolution filters and f be the filter size. For a single image, the convolution operation takes time:

$O(k\times f^2\times p)$

For n images, the overall time complexity of feature extraction becomes:

$O(n\times k\times f^2\times p)$

C. Classification Complexity

The fully connected layers classify features that have been extracted. If c is the number of output classes, the time for this classification is:

$O(n\times c)$

D. Overall System Complexity

Incorporating preprocessing, feature extraction, and classification, the time complexity in the proposed plant disease detection system combines all three stages. It can be expressed as:

$O(n \times k \times f^2 \times p)$

Despite the fact that models driven by convolutional neural networks (CNNs) are resource hungry, the applications of Graphics Processing Units (GPUs) help in decreasing both training and inference time considerably, thereby enhancing the applicability of the system in time-sensitive and extensive deployments in the field of agriculture.

VII. BLOCK DIAGRAM DESCRIPTION

Figure 2 shows the workflow of the plant disease detection system. Use case shows the data flow starting from the image input and ending at the result output.

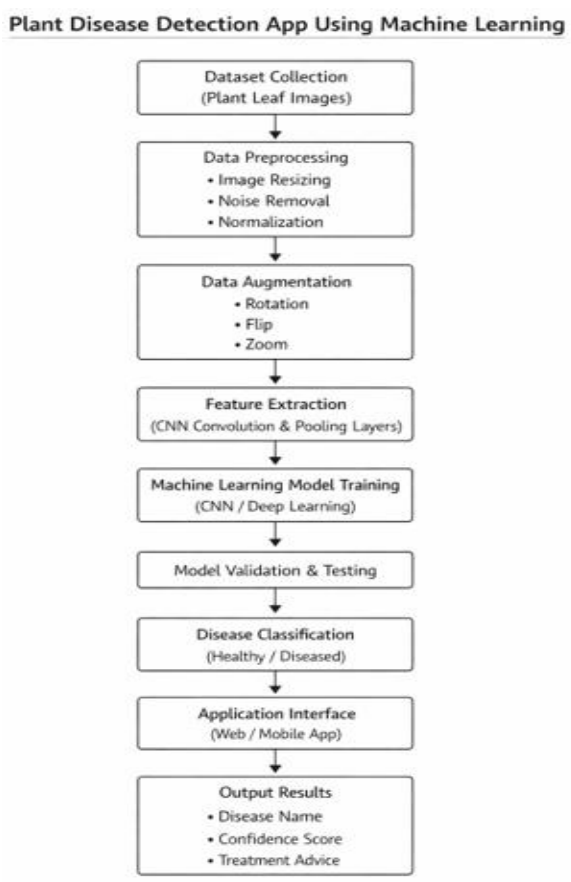

Fig. 2. Block diagram of the plant disease detection system

System starts with the input of leave images from the user and does image preprocessing. After that, it moves the pre-processed image into the CNN model for feature extraction and classification. Finally, the system presents the classification output to the user.

VIII. FLOWCHART OF PLANT DISEASE DETECTION

Process

The flowchart in Fig. ?? diagram we present of the put forth plant disease detection system. It details the step-by-step logical flow from image collection to the end which is disease diagnosis.

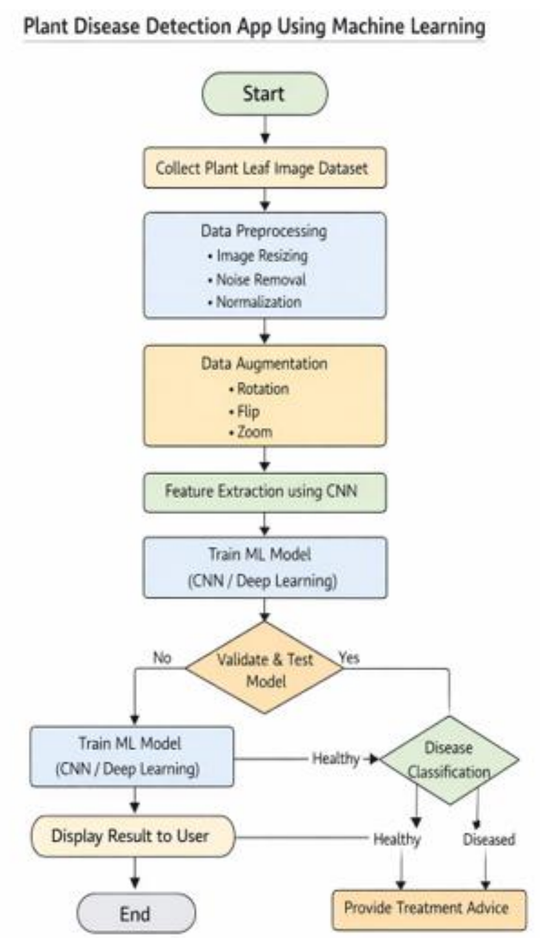

Fig. 3. Flowchart of the plant disease detection process

The process begins with the capture of a leaf image via a camera or mobile device. Once captured the image goes to preprocessing where it is resized, normalized, and augmented. These actions produce uniform input data which also see to better quality.

Post-processed images proceed to the Convolutional Neural Network (CNN) model for feature extraction and disease classification. The CNN analyses the leaf image to identify visual patterns and classify the leaf as healthy or diseased. Based on the classification, the system provides the user with the diagnosis via a web or mobile application. If the leaf is healthy, the system communicates a healthy status. If not, it specifies the disease. The process concludes after the final result is presented.

IX. EXPERIMENTAL RESULTS

The proposed plant disease detection system can be analysed with regard to system reliability and classification accuracy. In the case of the CNN, the data has to be divided into the training and validation sets. The validation set is used to assess performance after training the CNN on the training set to demonstrate the ability to generalize.

Here, classification accuracy is used, as it is the most common. Accuracy is the number of correctly classified images divided by the total number of images. To evaluate the CNN model more broadly, precision, recall, and the F1-score have been calculated as well. Each disease class has been assigned accuracy, precision, recall, and F1-score.

In fig: accuracy we see that the trend is of an increasing accuracy which we note as a steady improvement with each epoch that in turn proves the CNN's learning ability. Also the large gap between the training and validation accuracy indicates that the model is learning well without issue of overfitting.

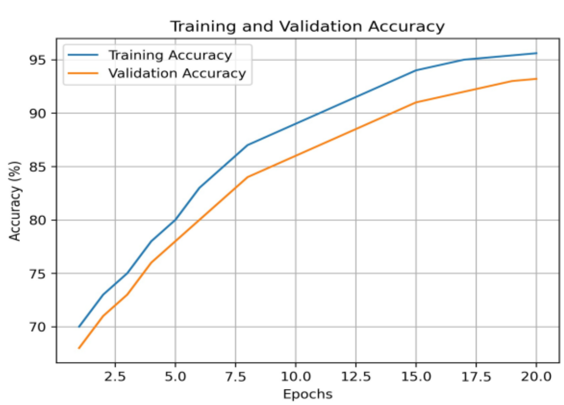

Fig. 4. Training and validation accuracy of the CNN model

The quantitative performance metrics achieved by the proposed system are summarized in Table eftab:metrics. The high accuracy and F1-score demonstrate the effectiveness of the CNN-based approach for plant disease detection.

Table I Performance Metrics

Metric	Value (%)
Accuracy	95.6
Precision	94.8
Recall	95.2
F1-Score	95.0

X. COMPARATIVE ANALYSIS

In this section we present a comparison of the put forth CNN based plant disease detection system with more traditional machine learning methods like Support Vector Machines (SVM) and Decision Trees. Traditional methods use hand crafted features which in turn limits their performance in various environmental settings. As opposed to this the CNN based approach is able to learn hierarchical features by itself. This results in improved accuracy and also greater robustness.

Table Ii Comparative Analysis Of Classification Techniques

Method	Feature Extraction	Accuracy (%)
Decision Tree	Manual	82.4
SVM	Manual	86.7
CNN (Proposed)	Automatic	95.6

The comparison shows that the suggested CNN-based method performs much better than traditional approaches when it comes to accuracy and reliability.

XI. RESULTS AND DISCUSSION

The experiment results show that our CNN-based plant disease detection system is effective in terms of classification accuracy and reliability. The CNN implementation offers benefits over traditional machine learning algorithms. It automatically learns features from the images, which leads to improved accuracy and confidence in our system. The results also indicate that deep learning methods fit well for agricultural applications, particularly for early disease detection.

Once trained, our model can be added to a web-based application. This allows users, such as farmers, to upload images of leaves and get real-time diagnoses for their plants.

The experimental results reveal that the CNN-based approach performs better than traditional machine learning techniques because it automatically learns key features from images. The high accuracy achieved shows how effective deep learning is for detecting plant diseases. The web-based application adds to accessibility and ease of use for farmers.

XII. CONCLUSION AND FUTURE WORK

This research has shown a plant disease detection system that uses machine learning and is based on Convolutional Neural Networks. The system proposed in this research can effectively detect plant diseases from the plant leaf images automatically and can classify the overall system with a high accuracy. The experiments' outcomes have proved that CNN can learn significant features in the planting system that can discriminate between healthy and unhealthy plants.

The proposed system can serve as an effective plant disease diagnosis system that provide a very practical and easy to use plant disease detection platform in the crop production, significantly reducing the application of fertilizer and chemical substances. The stable system can be scaled and integrated into the web system based on the modular frame, which is suitable for different farming environments.

In the future, the system proposed in this research can be expanded by using transfer learning and updating it with newer CNN models to improve accuracy and reduce training time. Deploying the system on mobile platforms can make it accessible for farmers in remote areas. Additionally, capturing plant images in the field in real-time and uploading them to the system through an IoT-based smart system can help monitor plants with less effort.

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

[4] J. G. A. Barbedo, "Factors influencing the use of deep learning for plant disease recognition," Biosystems Engineering, vol. 172, pp. 84, 91, 2018.

[5] A. Picon et al., "Crop disease classification using deep learning," Computers and Electronics in Agriculture, vol. 161, pp. 280, 290, 2019.

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

[7] Y. Lu et al., "Identification of rice diseases using CNNs," Neurocomputing, vol. 267, pp. 378, 384, 2017.

[8] S. Zhang et al., "Deep learning-based crop disease detection," IEEE Access, vol. 7, pp. 123453, 123463, 2019.

[9] A. Ramcharan et al., "Deep learning for cassava disease detection," Frontiers in Plant Science, vol. 8, 2017.

[10] E. C. Too et al., "Fine-tuning deep learning models for plant disease identification," Computers and Electronics in Agriculture, vol. 161, pp. 272, 279, 2019.

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep CNNs," NeurIPS, 2012.

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

[13] K. He et al., "Deep residual learning for image recognition," CVPR, 2016.

[14] C. Shorten and T. M. Khoshgoftaar, "Image data augmentation for deep learning," Journal of Big Data, 2019.

[15] K. G. Liakos et al., "Machine learning in agriculture: A review," Sensors, vol. 18, no. 8, 2018.

[16] A. Fuentes et al., "Real-time plant disease detection using deep learning," Sensors, vol. 17, no. 9, 2017.