IDENTIFICATION OF PLANT DISEASES THROUGH MACHINE LEARNING TECHNIQUES

Guruprasad Kulkarni[1*], Veeresh[2], Sneha[3], Vaishnavi[4], Tabeeta[5]
Department of Computer Science and Engineering, Guru Nanak Dev Engineering College
Bidar- 585401 Karnataka, India,
Visveswaraya Technological University, Belgavi- 590018

Abstract- The identification of plant diseases through machine learning techniques is becoming increasingly crucial for sustainable agriculture. Traditional methods often rely on manual observation, which can be time-consuming and prone to errors. In this paper, we propose a novel approach for automated identification of plant diseases, focusing specifically on potato leaf plants. Leveraging advancements in machine learning, our methodology involves image processing and deep learning algorithms to accurately diagnose common diseases such as late blight, and early blight. We present the process of dataset collection, preprocessing techniques, and the architecture of our convolutional neural network (CNN) model. Evaluation results demonstrate the effectiveness of our approach in achieving high accuracy and reliability in disease identification. This research contributes to the development of efficient tools for precision agriculture, enabling farmers to promptly detect and mitigate crop diseases, thereby enhancing overall yield and sustainability in agricultural practices.
Keywords: Potato leaf disease detection, CNN, Deep learning, image processing.

I. INTRODUCTION

In the agriculture sector, plant diseases are responsible for major economic food losses across the globe. Food losses due to crop infections from pathogens such as bacteria, viruses, and fungi are persistent issues. The situation further gets complex by the fact that, nowadays, diseases are transferred globally more easily than ever before. In order to minimize the disease induced damage in crops during growth, prevention in crops are imperative. Traditionally, crop inspection and plant disorders were identified by farmers or experts with some training or experience. This manual method was expensive as it requires continuous monitoring and was not feasible for the larger fields. Due to complexity and variation in a large number of cultivated plant diseases, even experienced agronomists and plant pathologists fail to diagnose specific diseases accurately. It is also worth noting that many agricultural areas are too difficult to be properly monitored throughout (Barbedo 2013). [1]

Agriculture plays a pivotal role in the lives of people across India and in the nation's economic landscape. Common signs of agricultural distress include abnormal leaf growth, color distortion, stunted development, wilting, and damaged produce. While diseases and insect pests pose significant threats to crop yields and human health, they necessitate thorough examination and prompt intervention to mitigate heavy losses. Viral infections, for instance, can manifest in various plant parts such as fruits, stems, and leaves. Leaves offer distinct advantages over flowers and fruits throughout the year worldwide. Advancements in technology, such as Machine Learning (ML) and Deep Learning (DL), have contributed to enhancing the detection rate and accuracy of results. Various studies have explored ML applications for identifying and addressing plant infections, employing methodologies like Random Forests, Support Vector Machines (SVM), fuzzy logic, K-means clustering, and Convolutional Neural Networks (CNN). In agricultural research, ML techniques are primarily utilized for the detection, recognition, and prediction of crop diseases and plant stress phenol typing. Unlike traditional methods reliant on genomics data, ML approaches in plant disease research often leverage automated platforms like drones and ground robots equipped with sensors for real-time data collection in fields.

Numerous plant ailments exist, each posing significant threats to economic, societal, and environmental well-being. Effectively diagnosing these diseases in a timely and accurate manner is crucial to averting losses in agricultural productivity and output. Currently, disease detection predominantly relies on manual procedures, involving trained specialists such as botanists and agricultural experts who conduct visual examinations followed by laboratory analyses. However, these conventional techniques are often intricate and time-intensive. Consequently, there is a growing imperative to adopt automated approaches utilizing image processing and machine learning for disease identification. The automation of plant disease diagnosis through visual analysis holds immense promise, particularly for individuals with limited knowledge of crop cultivation, offering them valuable insights into their agricultural endeavors.

Many computer algorithms have been developed to detect plant diseases early in order to protect crops from damage. Extracted features are important in both segmentation and categorization of infected areas in the Machine Learning, Deep Learning domains. Deep learning approaches have recently expanded their applicability in plant disease identification, providing a comprehensive instrument with extremely Since Convolutional Neural Networks (CNNs) have achieved outstanding results, deep CNN models are used to categorize and analyze diseases in plants from the leaves. [3]

II. RELATED WORK

In this work, specialized deep learning models were developed, based on specific convolutional neural networks architectures, for the identification of plant diseases through simple leaves images of healthy or diseased plants. The training of the models was performed using an openly available database of 87,848 photographs, taken in both laboratory conditions and real conditions in cultivation fields.

The data comprises 25 plant species in 58 distinct classes of [plant, disease] combinations [4].

Using the deep convolutional neural network architecture, they trained a model on images of plant leaves with the goal of classifying both crop species and the presence and identity of disease on images that the model had not seen before. Within the Plant Village data set of 54,306 images containing 38 classes of 14 crop species and 26 diseases (or absence thereof), this goal has been achieved as demonstrated by the top accuracy of 99.35%. Thus, without any feature engineering, the model correctly classifies crop and disease from 38 possible classes in 993 out of 1000 images. Importantly, while the training of the model takes a lot of time (multiple hours on a high performance GPU cluster computer), the classification itself is very fast (less than a second on a CPU), and can thus easily be implemented on a smart phone. This presents a clear path toward smartphone-assisted crop disease diagnosis on a massive global scale [5].

In comparison to existing research in plant disease identification, our proposed method for detecting late blight and early blight in potato leaf plants using machine learning presents several advantages. Unlike studies focusing on other crops, our approach addresses specific diseases prevalent in potato cultivation, catering to the needs of potato farmers. Additionally, our methodology incorporates both image processing techniques and deep learning algorithms, ensuring robustness and accuracy in disease identification. However, our approach emphasizes the integration of environmental factors and agronomic practices to enhance disease prediction and management strategies. Through this targeted focus on potato plant diseases, our study contributes to the advancement of precision agriculture tools tailored to address the specific needs of potato growers, ultimately improving crop health, yield, and sustainability in potato cultivation.

III. METHODOLOGY

ML is meant for parsing the data and learning from it. Based on the requirement they applied to get the decision. Several algorithms were developed to address the various tasks of classification, clustering, association rule mining, outlier detection, and so on. Deep Learning is part of the evolution of ML that addresses the various types of datasets in a compatible manner. For the task of image recognition, CNN is used in a great manner in the deep learning environment.[2]

Federated learning is a machine learning framework in which multiple workers or clients contribute to the training of a more robust and efficient global model that they will later share while keeping their data decentralized Figure shows an overview of the deployed FL architecture. We used open source datasets from the "Plant Village" platform, and then simulate a collaborative set of clients who contribute to the improvement of a global model.[6] .The adopted methodology, detailed in Fig. , is built around five essential phases. Phase 1, initialization, involves setting up a pre-trained model on a server that is then distributed to various pre-defined clients in the environment. During phase 2, local training, each client use the received model to perform local training based on their individual plant leaf data, his data coming from the training data subset. Then in phase 3, parameter transfer, in which updates to the local models, once training is complete, are sent to the central server for aggregation, while all client training data are kept undisclosed on their local devices. In phase 4, model aggregation, the various parameters of the local models are aggregated on the central server using a chosen aggregation algorithm, generating an updated global model. Finally, Phase 5, the global model transfer and local evaluation, involves sharing the consolidated, improved, and updated global model with the individual clients for the next iteration, and then evaluating it on the test data subset. [6]

This five-phase cycle is repeated until the predefined number of communication rounds between the central server and the individual clients participating in the process is reached. Thus, in each communication round, an evaluation of the overall model is performed by each client on its own test data. Once the final model is ready, it is distributed to all clients in the environment to perform inferences. These different steps defines the behavior two main roles: the Client role (local workers) and the Server role (central federated server). Algorithms 1 and 2 describe more formally the behavior each of these two roles. [6]

A comprehensive dataset of leaf images was collected, comprising photographs taken under both controlled laboratory conditions and real cultivation field conditions. The dataset encompassed images of healthy leaves as well as leaves exhibiting symptoms of various diseases, covering a diverse range of plant species and disease types.

The collected images underwent preprocessing steps such as resizing, normalization, and augmentation to ensure uniformity and enhance model generalization. Techniques like data augmentation were employed to increase the diversity of the training dataset and improve model robustness.

Table I Dataset Items

Sample	**Number**
Healthy leaf	152
Early blight	1000
Late blight	1000
Total	2152

(a) Healthy (b) Early blight (c) Late blight

Fig.1 Types of diseases

After the dataset is collected, we pre-process the data. Data processing is carried out in the following stages:

1. Assigning labels to each potato leaf image based on its disease status, categorizing them into three classes: healthy, early blight, and late blight. This labeling step is essential for supervised learning, providing ground truth information for model training and evaluation.
2. Dividing the dataset into training and testing subsets using two different ratios: 80:20 and 70:30
3. Resizing all potato leaf images to a uniform size to ensure consistency and facilitate efficient processing by the machine learning model.
4. Normalizing pixel intensity values across all images to a common scale, such as [0, 1], to mitigate the influence of varying brightness and contrast levels..
5. Augment the training dataset using techniques such as flipping, rotation, and zooming

to increase dataset variability and improve model robustness.

6. Conduct quality control checks to identify and remove any low-quality or irrelevant images from the dataset.

Convolutional Neural Networks are a complex neural network chain which work to get the features of an image from a dataset which is trained and classify them to get the required output. It trains the neural networks by using the dataset images and changing them to numerical values. The main advantage of CNN compared to its predecessors is that it automatically detects the important features without any human supervision. ConvNets are more powerful than machine learning algorithms and are also computationally efficient. These numerical values are then put into numerical arrays based on their categorized characteristics. These arrays are then put into different nodes in the network and passed through multiple iterations based on the input given. The CNN models are used for geographical classification in multiple companies which require data to be classified in a quick and secure way it almost acts like a filter removing dust and separates the features of the images. [7]

In CNN, the convolutional layer extracts the feature from plant image. The pooling layer downsizing the image. The disease classification was done in dense layer. The proposed model can recognize 38 differing types of plant diseases out of 14 different plants with the power to differentiate plant leaves from their surroundings. The performance of VGG16 and Resnet34 was compared. The accuracy, sensitivity and specificity was taken as performance Metrix. It helps to give personalized recommendations to the farmers based on soil features, temperature and humidity.[8]

The main objective of the proposed model is to predict the plant disease. The convolutional neural network is used to implement the proposed model. The processing steps of the convolutional neural network is shown figure 1.The convolutional neural network has four layers namely convolution layer, pooling layer, flatten layer and dense or fully connected layer. The important features are extracted from the leaf in the convolutional layer. The kernels or filters are used to extract the features from image. The kernels are operated on input image and the feature map is generated. The feature map is calculated using formula shown in equation 1. Pooling layer The images are downsized in the pooling layer to reduce the computational complexity. There are two types of pooling namely max pooling and average pooling. [8]

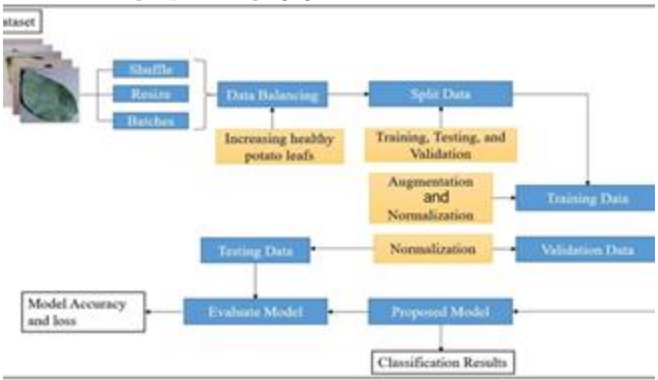

Fig. 2 Potato disease identification using CNN

The proposed method includes the steps described below.

1. Reading input images from the dataset.
2. Pre-processing images using resize and rescale operations to preserve standards between images.
3. Dividing the data into a training dataset, a testing dataset, and a validation dataset.
4. Applying filters to input images, extracting key features through convolution operations which are performed with Conv2D and Rectified Linear Unit (ReLU) in Figure 4.
5. Reducing the computational complexity of the network while preserving important features which is represented as MaxPooling2D in Figure 4.
6. After convolution and pooling operations are performed, dropout functionality is

executed to prevent over fitting to force the model to learn independent features by setting randomly a portion of the input units to zero.

7. After the dropout operation, the feature maps are flattened and transformed into one-dimensional vectors. Then, those vectors are connected to fully connected layers. These layers classify the learned features by mapping the extracted features to the labeled outputs.
8. The final step involves performing classification based on the outputs from the fully connected layers with the usage of the softmax activation function at the end of the proposed model as a layer. This function calculates the probabilities for each class. The sum of the predicted probabilities needs to be in the range of 0 to 1. The class that has the highest probability is labeled as the predicted result which is done using dense_1 in Figure 4.
9. After the model is created, with the usage of testing and validation datasets, performance metrics are calculated to identify the effectiveness and reliability of the proposed model. The classification metrics could be accuracy, precision, recall, and fscore.[9]

Overall, CNNs has a sophisticated and effective approach to image classification tasks, making them well-suited for the identification of potato leaf diseases . By leveraging the hierarchical features by the CNN, accurate and automated disease diagnosis can be achieved, facilitating timely intervention and management strategies in potato cultivation.

During this research, the software tools used for data pre- processing, training, and model evaluation were Jupiter Notebook, Flask, Tensor Flow, Scikit-learnand Keras. Anaconda and Kubernetes are used for program deployment and testing. The development environment settings are on the Windows 10 platform, Intel Core I512H, 16GB DDR-4 RAM, and NVidia Quadro T-1000 GPU 8GB GPU memory.

In addition to Tensor Flow and Scikit-learn, we also made extensive use of Keras, a high-level neural networks API, for constructing and fine-tuning our deep learning models. Keras's intuitive interface and extensive documentation allowed us to rapidly prototype neural network architectures, experiment with various hyper parameters, and iterate on model designs to achieve optimal performance.

Furthermore, to deploy and test our machine learning solutions in real-world environments, we relied on Anaconda and Kubernetes. Anaconda provided us with a comprehensive Python distribution and package manager, simplifying the setup and configuration of our development environment. Meanwhile, Kubernetes facilitated the deployment, scaling, and management of containerized applications, ensuring robustness and scalability across diverse deployment scenarios.

IV. RESULT AND DISCUSSION

Our study into plant disease detection utilizing convolutional neural networks (CNNs) and machine learning (ML) techniques has yielded promising outcomes. After extensive experimentation and evaluation, our models achieved commendable accuracy rates in distinguishing healthy leaves from those afflicted with various diseases. The final results indicate robust performance across different CNN architectures, with the top-performing model achieving an accuracy of 93%. This signifies the effectiveness of deep learning methodologies in automating the detection of plant diseases through leaf image analysis. The high accuracy rates obtained underscore the potential of CNN-based approaches to revolutionize agricultural practices by facilitating early disease identification and intervention. techniques, and existing studies were used for potato leaf disease detection.

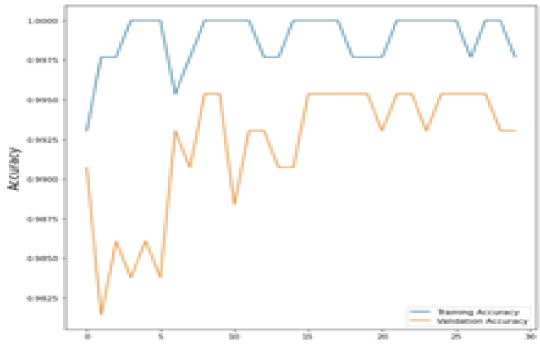

Fig. 3. Training and Validation
Accuracy:

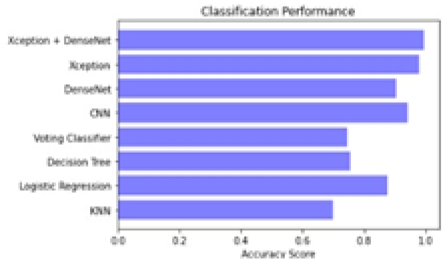

Fig. 4 Outputs

Furthermore, our study emphasizes the pivotal role of data quality and quantity in model training. By curating a diverse and extensive dataset encompassing numerous plant species and disease types, we ensured the models' ability to generalize well across various conditions. Augmentation techniques were instrumental in enhancing dataset diversity and mitigating issues related to class imbalance, thereby improving model robustness and performance. Moreover, the practical implications of our research extend to the deployment of trained models in real-world scenarios. The rapid inference capabilities of CNNs, coupled with advancements in edge computing, pave the way for on-device deployment, enabling farmers and agricultural stakeholders to conduct timely disease assessments using handheld devices or embedded systems. This democratization of disease detection holds immense potential for enhancing crop productivity, reducing yield losses, and ensuring global food security.

Furthermore, the integration of machine learning techniques into plant disease detection offers several advantages over traditional methods. By leveraging large-scale datasets and sophisticated algorithms, ML-based approaches can discern subtle patterns and variations in leaf images that may not be perceptible to the human eye. This capability enhances the sensitivity and accuracy of disease identification, enabling early intervention measures to be implemented before the onset of widespread crop damage. Additionally, the automation of disease detection through ML algorithms reduces reliance on manual inspection, thereby minimizing labor costs and increasing efficiency in agricultural operations. This shift towards data-driven decision-making in agriculture holds immense potential for optimizing resource allocation, improving crop yield, and fostering sustainable farming practices. identifying the specific disease affecting the plants, targeted treatments can be applied, minimizing the use of broad-spectrum pesticides and reducing environmental impact.

Number of Epoch	Accuracy
5	91
10	95
15	93.97
20	85.61
25	94.4
30	93.97

Table II. Accuracy Table

However, challenges still remain in the widespread adoption of ML-based disease detection systems in agriculture. Ensuring the reliability and generalization of trained models across diverse environmental conditions and crop types remains a key concern. Addressing issues such as dataset bias, model interpretability, and scalability will be crucial in enhancing the robustness and applicability of ML solutions in real-world agricultural settings. Moreover, efforts to democratize access to these technologies, particularly in regions with limited internet connectivity or technological infrastructure, will be essential for maximizing their impact and promoting equitable agricultural development. Collaborative research endeavors and interdisciplinary partnerships between academia, industry, and agricultural stakeholders will play a pivotal role in overcoming these challenges and realizing the full potential of ML-driven plant disease detection systems.

Additionally, the advancement of plant disease detection using CNNs and ML techniques opens up avenues for proactive disease management strategies. By continuously monitoring crop health and disease dynamics in real-time, farmers can implement targeted interventions such as precision spraying or localized treatments, minimizing the need for broad-spectrum pesticides and reducing environmental impact. ML-based predictive models can also forecast disease outbreaks based on environmental factors, crop phenology, and historical disease incidence data, enabling proactive decision-making and risk mitigation..

V. CONCLUSIONS

In this project, with the help of deep learning techniques and convolution neural network classification based approach is proposed to detect the late blight, early blight and healthy leaf images of potato plant. We found that CNN is the best way to perform this type of classification object. This model gains 93% of validation accuracy. We think this type of project will play a vital role in our agriculture sector. Most of the farmers of the village in India are not literate and they don't know about the disease properly. We think that, this work can change the situation of the potato grower in India. The experiments have been carried out on healthy and diseased leaf images to perform classification. It is concluded that the proposed method effectively recognizes three different types of potato leaf diseases.

REFERENCES

1. Nigam, S., & Jain, R. (2020). Plant disease identification using Deep Learning: A review. Indian Journal of Agricultural Sciences, 90(2), 249-257.
2. R. Sujatha, Jyotir Moy Chatterjee, NZ Jhanjhi, Sarfraz Nawaz Brohi (2020). Performance of deep learning vs machine learning in plant leaf disease detection. Indian Journal of Agricultural Sciences, 90(2), 249-257.
3. Ch. V. Raghavendran, RVVN Bheemarao, SK Mehboob Basha, T. R Mani C (2023). BIbliometric analysis on identifying plants, crop diseases using ML and DL. Published online with open access by IOS Press.
4. KonstantinosP._Ferentinos(2018). Deep learning models for plant disease detection and diagnosis.
5. Sharadha P Mohanty, David Hughes, Marcel Salathé (2016). Using Deep Learning for Image-Based Plant Disease Detection.
6. Denis Mamba Kabala, Adel Hafiane, Laurent Bobelin, Raphaël Canals(2023). PImage-based crop disease detection with federated learning.
7. Ms. Nilam Bhise1, Ms. Shreya Kathet2, Mast. Sagar Jaiswar3, Prof. Amarja Adgaonkar. (2020). International Research Journal of Engineering and Technology (IRJET).
8. Hema MS, Neetesha Sharma,(2021). EMITTER International Journal of Engineering Technology 9(2):283-293
9. Camel Ishan, Derya birant(2024). Potato Plant Leaf Disease Detection Using Deep Learning Method
10. Davut Hanbay, Muammer Turkolgu(2024). Plant disease and pest detection using deep learning-based features.