RESEARCH ARTICLE OPEN ACCESS

Plant Disease Diagnosis Using Convolutional Neural Network

Divyansh Saxena[1], Tejas Rai[2]

Under Guidance of Associate Professor Miss Jyoti Agarwal

1(Computer Science Engineering, Graphic Era (Deemed to be University), Dehradun
Email: divyanshssaxena1@gmail.com)

2 (Computer Science Engineering, Graphic Era (Deemed to be University), Dehradun
Email: raitejas16@gmail.com)

Abstract:

Plant disease prediction system is an essential application of machine learning in the field of agriculture. In this system, convolutional neural networks (CNNs) are used for predicting plant diseases based on the input image of a plant leaf. The proposed system has five layers of CNN, which have been trained on different plant datasets to improve the prediction accuracy. The system has been trained on a large dataset of plant images to identify the symptoms of different plant diseases accurately. The CNN model uses image processing techniques to extract features from the input image and classify it into the corresponding plant disease category. The system has achieved an accuracy of 91%, which is a promising result for plant disease prediction. The system can help farmers in the early detection and prevention of plant diseases, leading to better crop management and higher yield. The proposed system can also reduce the use of chemical pesticides, which can be harmful to the environment and human health. Overall, the plant disease prediction system using machine learning with six layers of CNN is a significant step towards sustainable agriculture.

Keywords — **machine learning, convolutional neural network, image processing, dataset, feature extraction.**

I. INTRODUCTION

This document is a template. An electronic copy can be downloaded from the conference website. For questions on paper guidelines, please contact the conference publications committee as indicated on the conference website. Information about final paper submission is available from the conference website. An estimated 1.1 billion people are engaged in agriculture and agriculture related works which accounts to 28% of the global employment in 2013. Agriculture sector contributes around 6.4% to world's total economic production (GDP). Agriculture plays a very important role in transforming economies to reach the upper heights and become the developed nations. It is still one of the main occupations in many countries and the GDP of many countries still heavily relies on farming and raising livestock especially of developing countries. In India, agriculture and allied sector accounted for 21% of GDP [1]. India holds a prominent position in global agriculture as it is the second largest producer of essential food staples such as wheat and rice. Additionally, India's agricultural prowess extends to the production of several other crops including dry fruits, textile raw materials, pulses, root crops, farmed fish, eggs, coconut, sugarcane, and a diverse range of vegetables. To export and feed billions of people and to keep them healthy, the quality of food which is being given to them matters too. People should be given the right type of food devoid of any ailments or pathogens. But as in humans, plants too suffer from diseases which disrupts their growth and affects their quality, thus affecting the health of humans too.

Crop diseases are a major threat to all the ones consuming it. A lot is at stake when a farmer's crop is struck with any unwanted thing that can cause significant loss in production, economic losses, and a reduction in the quality and quantity of agricultural products. Plant infections also affects humans by secreting toxic enzymes. A plant's growth mainly depends on these four 2 factors –

light, temperature, water, and nutrients. Wrong fluctuation in any of these can cause serious harm to the crop. Climate change and harm from locust swarms is also a major cause in plant infections and devastations. In recent years, there is also an increase in bacterial, fungal, and viral infections which can affect a plant at different stages of its growth. Diagnosis of plant diseases is very crucial for the plants and humans. The early detection of a disease has a positive effect on a plant's health. Manual prediction can be slow and time consuming so here's where technology comes into play. Machine learning and deep learning techniques are now being increasingly used to predict the health and future growth of a crop. Computer vision has emerged as one of the leading cutting-edge technologies in the past decade and has made its presence in all fields and domains and agriculture is not untouched. The integration of AI systems in agriculture has led to the advancement of precision agriculture, which is focused on enhancing the quality and accuracy of harvests. By leveraging AI technology, farmers are now able to identify and address various issues such as plant diseases, pest infestations, and inadequate nutrition on their farms. Convolutional Neural Networks (CNNs) have led to significant progress in this using image processing techniques. Many studies and papers have been published in this domain to explore the never-ending vicinity. Many ML models are being used to predict the future of agriculture and its influence on us but the thing which matters the most is the accuracy of prediction.

II. LITERATURE REVIEW

The recent trends in CNN and deep learning architectures in agricultural application are discussed. Plant disease diagnosis through optical observation has a significantly higher degree of complexity. The infections and complexities are analysed using CNN and image processing techniques in tomatoes, potatoes, and bell pepper [2]. Similar approach to detect a greater number of diseases of multiple crops is done by others [3]. A new approach is provided by researchers which features CNN along with Learning Vector Quantization (LVQ) algorithm. Colour information is actively used for plant leaf detection is this method [4]. An explanation of DL models used to visualize different plant diseases and several research gaps are identified to obtain greater transparency for detecting diseases in plants [5].

The feasibility of CNN for disease prediction from leaf images are taken under the natural instances. The model is designed based on the LeNet architecture [6]. The employed neural networks models are used to aid in the classification of the input image into respective disease classes. This proposed system has achieved an average accuracy of 94-95 % [7]. The different approaches to treat aspects related to disease detection, characteristics of the dataset, the crops and pathogens investigated [8]. The dataset from Plant Village from 48 different classes and got an accuracy of 95.81% using different CNN hyperparameters [9]. The CNN Model with Optimized Activation Function is used with TensorFlow framework. Further area affected by disease is calculated by using K - means clustering algorithm for optimization of fertilizer usage [10]. This CNN model consumes 3.8 seconds for identifying the image class with more than 94.5% accuracy [11]. A hybrid model based on Convolutional Autoencoder (CAE) network and Convolutional Neural Network (CNN) achieves 99.35% accuracy [12]. Two unique models namely SCNN-KSVM (Shallow CNN with Kernel SVM) and SCNN-RF (Shallow CNN with Random Forest) outperform other deep models on the indicators of precision, recall, and F1-score [13]. A simpler approach utilizing the leaves images and then segmenting and extracting its features [14]. A group concluded that S-CNN model improves significantly over F-CNN model with accuracy of 98.6% [15].

A new model achieved a high accuracy of 96.67% for identifying the diseases in mango plant, hence showing the feasibility of its future usage in real time applications [16]. In another paper the accuracy increased to 98.3% of earlier models by changing image brightness by a random value of a random width of image after image augmentation [17]. Shallow VGG with Xgboost model which outperforms different deep learning models in terms of accuracy, precision, recall, f1-score, and specificity [18]. The researchers have developed

CNN models using simple leaves images of healthy and diseased plants, through deep learning methodologies containing 25 different plants in a set of 58 distinct classes of plant-disease combinations and got an accuracy of 99.53% [19]. The three main Architecture of the Neural Network: Faster Region-based Convolution Neural Network (Faster R-CNN), Region-based Fully CNN(RCNN) and Single shot Multibook Detector (SSD) are used to achieve an accuracy of 94.6% [20].

III. METHODOLOGY

Plant diseases have a significant impact on the agricultural industry and can lead to significant economic losses. Early detection of plant diseases can help prevent their spread and minimize crop losses. Machine learning algorithms, particularly Convolutional Neural Networks (CNN), have shown promising results in predicting plant diseases from images of plant leaves. In this paper, we propose a plant disease prediction system using machine learning with a focus on a six-layer CNN model.

The dataset used in this research comprises images of different plants affected by various diseases. The model is trained using backpropagation, and the weights are updated using the Adam optimization algorithm. The model's hyperparameters are tuned using a validation dataset to achieve the best performance. The model is then tested using the testing dataset to evaluate its performance in predicting plant diseases. Our methodology involves training a six-layer CNN model using the training dataset, tuning the hyperparameters using a validation dataset, and evaluating the model's performance using the testing dataset. The paper should also discuss the limitations and future scope of the proposed system.

A. Dataset

For this plant disease prediction project using machine learning, a dataset containing 38 classes of healthy and unhealthy plants such as corn, grape, apple, tomato, potato, and others is utilized. The dataset includes a large number of images of plants, with each image labelled with the name of the plant species and the specific disease (if any) affecting it. The dataset was carefully selected to ensure that it includes a diverse range of images of both healthy and diseased plants, with a sufficient number of samples per class to avoid overfitting. By utilizing this dataset, we trained a machine learning algorithm to accurately predict whether a new plant is healthy or diseased, and if diseased, which specific disease is affecting it.

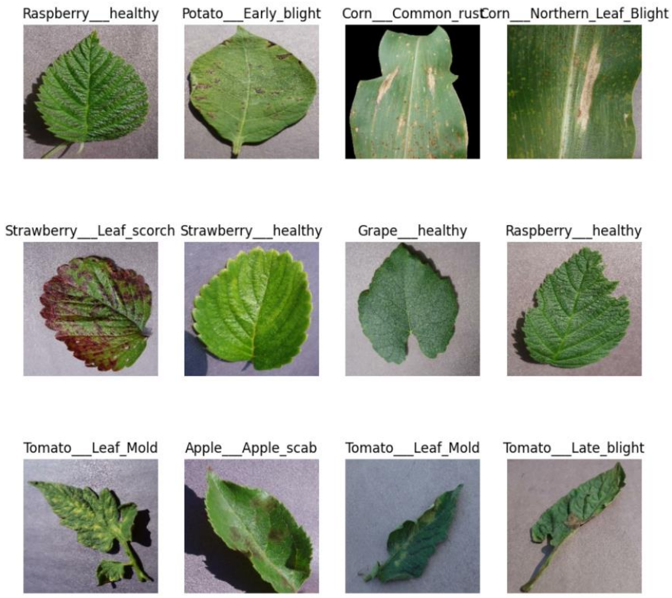

Fig. 1 Sample images of different plants from the dataset

B. Data pre-processing and augmentation

Data pre-processing and augmentation are crucial steps in building a robust and accurate system. Data pre-processing involves cleaning and preparing the dataset before training the machine learning model. It includes tasks such as data cleaning, data normalization, and feature engineering. Data cleaning is the process of removing any irrelevant or erroneous data points from the dataset. Data normalization is the process of scaling the features of the dataset to the same range. Feature engineering is the process of creating new features from the existing ones that can help the model to better learn the patterns in the data.

Data augmentation involves creating new data points by applying transformations to the existing data. This is a powerful technique for increasing the size of the dataset, which can help to prevent overfitting. Image transformations include flipping, rotating, zooming, and changing the brightness and contrast of the image.

In the context of the Plant Village dataset, the images have a fixed size of 256 x 256 pixels. To prepare the dataset for machine learning, data processing and image augmentation techniques are applied using the Keras deep-learning framework. Keras is a powerful and user-friendly deep-learning library that provides a high-level API for building and training neural networks.

TABLE I
MODEL SUMMARY

Layer (Type)	Output Shape	Param #
Sequential (Sequential)	(32, 256, 256, 3)	0
Sequential_1 (Sequential)	(None, 256, 256, 3)	0
conv2d (Conv2d)	(None, 254, 254, 32)	896
max_pooling2d (MaxPooling2D)	(None, 127, 127, 32)	0
conv2d_1 (Conv2d)	(None, 1254, 1254, 64)	18496
max_pooling2d_1 (MaxPooling2D)	(None, 62, 62, 64)	0
conv2d_2 (Conv2d)	(None, 60, 60, 64)	36928
max_pooling2d_2 (MaxPooling2D)	(None, 30, 30, 64)	0
conv2d_3 (Conv2d)	(None, 28, 28, 64)	36928

C. Feature extraction and selection

Feature extraction involves extracting relevant features from the raw data, while feature selection involves selecting a subset of the available features that are most informative for the prediction task. feature extraction can be done using various techniques, such as deep learning, texture analysis, and shape analysis. Deep learning models, such as convolutional neural networks (CNNs), can automatically learn relevant features from the raw image data without the need for manual feature extraction. These learned features can be used as input to a classification model to predict the disease class. Once the features have been extracted, feature selection can be used to identify the most important features for the prediction task. There are several methods for feature selection, such as filter methods, wrapper methods, and embedded methods.

D. Convolutional Neural Networks Layers

In this project, a Convolutional Neural Network (CNN) with 6 layers is used, where the first 5 layers use the Rectified Linear Unit (ReLU) activation function and the last one is using the Softmax activation function. The purpose of the project was to develop a model that can accurately identify the type of disease affecting a plant by analysing its image. The CNN architecture was chosen due to its effectiveness in image recognition tasks.

The first layer of the CNN used was a convolutional layer that applied a set of filters to the input image, extracting features that were relevant to the task at hand. The next layer was a pooling layer that reduced the dimensionality of the feature maps, helping to reduce overfitting and improve the computational efficiency of the model. The subsequent layers followed the same pattern, with convolutional layers extracting increasingly complex features and pooling layers reducing their dimensionality.

The last layer of the CNN was a fully connected layer that mapped the output of the previous layers to a probability distribution over the different classes of plant diseases. This layer used the Softmax activation function, which ensured that the sum of the probabilities for each class was equal to one. During the training phase, the CNN was fed a large dataset of labelled images of plants with different types of diseases. By adjusting the weights of the network through backpropagation, the CNN learned to accurately classify the images into their respective classes. Overall, the CNN architecture with 6 layers and ReLU activation in the first 5 layers and Softmax activation in the last layer proved to be a highly effective approach to solving the plant disease prediction problem.

E. Result

The model developed in this project achieved an accuracy of 91% on the test dataset, which is a strong indicator of its effectiveness. The model was also able to avoid overfitting, which is a common challenge in deep learning models. This means that the model was able to generalize well to new data, making it suitable for use in real-world applications.

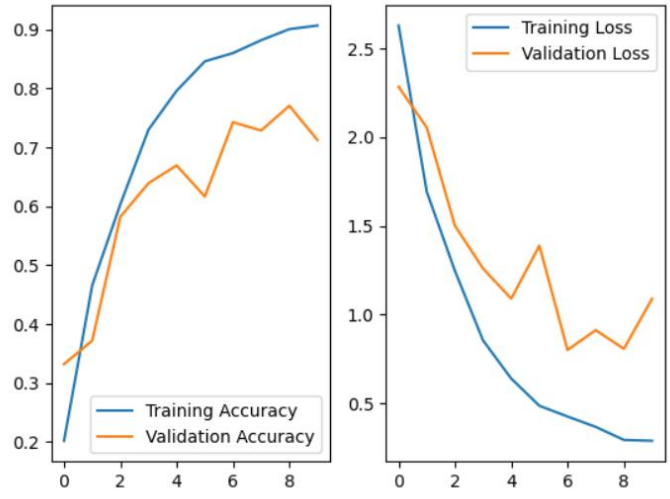

Fig. 2 Training and Validation graph of the Model

F. Future Challenges

Despite the significant progress that has been made in the field of plant disease prediction using machine learning, there are still several challenges that need to be addressed in order to improve the accuracy and effectiveness of these systems. One of the biggest challenges in developing accurate machine learning models for plant disease prediction is the availability and quality of data. Many crops and diseases lack the necessary data to develop effective models. Machine learning models developed for plant disease prediction may not be transferable to new locations, environments, and growing conditions. Another challenge is the need for rapid detection and response to plant diseases. If the diagnosis and treatment of plant diseases are delayed, it can lead to significant crop losses and economic damage. The development and deployment of machine learning models for plant disease prediction can be expensive. Therefore, there is a need for cost-effective solutions that can be easily deployed and maintained by farmers.

G. Future Scope

The future scope for plant disease prediction system using machine learning is promising, as there are several areas where machine learning can be further applied and improved upon. Machine learning models can be integrated with Internet of Things (IoT) devices, such as sensors and drones, to collect real-time data. Cloud-based solutions can be developed to provide access to machine learning models and data analytics tools to farmers in remote and isolated areas. Machine learning models can be used to automate the diagnosis and treatment of plant diseases, potentially reducing the need for human intervention. We can improve the accuracy and effectiveness of these systems, leading to significant benefits for farmers and the agriculture industry.

IV. CONCLUSION

In conclusion, the use of machine learning and convolutional neural networks (CNN) has shown great potential in accurately predicting plant diseases. The accuracy rate of 91% achieved in this system is an indication of the effectiveness of this technology in addressing the challenges faced in the agricultural sector. The use of machine learning and CNN in plant disease prediction has the potential to revolutionize the agriculture industry by improving crop yields and reducing the use of pesticides. As the technology continues to evolve, we can expect even more accurate and efficient models to be developed, further increasing the effectiveness of plant disease prediction and prevention.

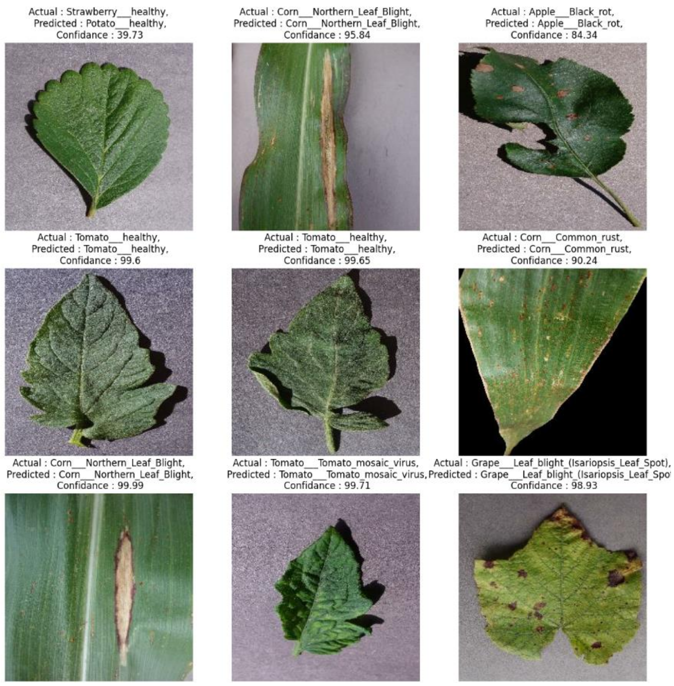

Fig.3 Accurate prediction of different plants using our model.

REFERENCES

1. *Sector-wise GDP OF India [Online]: Available: https://statisticstimes.com/economy/country/india-gdp-sectorwise.php*

2. *G. Shrestha, Deepsikha, M. Das and N. Dey, "Plant Disease Detection Using CNN," 2020 IEEE Applied Signal Processing Conference (ASPCON), pp. 109-113, Oct 2020.*

3. *U. Shruthi, V. Nagaveni, and B. K. Raghavendra. "A review on machine learning classification techniques for plant disease detection.", 5th International conference on advanced computing & communication systems (ICACCS), IEEE, pp. 281-284, March 2019.*

4. *M. Sardogan, A. Tuncer, and Y. Ozen. "Plant leaf disease detection and classification based on CNN with LVQ algorithm.", 3rd international conference on computer science and engineering (UBMK), IEEE, pp. 382-385, Sept 2018.*

5. *S. M. Hammad, J. Potgieter, and K. M. Arif, "Plant Disease Detection and Classification by Deep Learning" Plants, vol. 8, no. 11, pp. 468-470, Oct 2019.*

6. *S. Wallelign, M. Polceanu and Cédric Buche. "Soybean plant disease identification using convolutional neural network.", FLAIRS conference, pp. 146-151, May 2018*

7. *P. Tm, A. Pranathi, K. SaiAshritha, N. B. Chittaragi and S. G. Koolagudi. "Tomato leaf disease detection using convolutional neural networks.", Eleventh international conference on contemporary computing (IC3), IEEE, pp. 1-5, August 2018.*

8. *A. Abade, P. A. Ferreira and F. B. Vidal. "Plant diseases recognition on images using convolutional neural networks: A systematic review.", Computers and Electronics in Agriculture, vol. 185, pp. 106125-106134, June 2021.*

9. *J. Trivedi, Y. Shamnani and R. Gajjar. "Plant leaf disease detection using machine learning.", International conference on emerging technology trends in electronics communication and networking, Springer, vol. 1214, pp. 267-276, Feb 2020.*

10. *S. Y. Yadhav, T. Senthilkumar, S. Jayanthy and J. A. Kovilpillai. "Plant disease detection and classification using cnn model with optimized activation function.", International Conference on Electronics and Sustainable Communication Systems (ICESC), IEEE, pp. 564-569, July 2020*

11. *P. Deepalakshmi, K. Lavanya, and P. N. Srinivasu. "Plant leaf disease detection using CNN algorithm." International Journal of Information System Modeling and Design (IJISMD), vol. 12, no. 1, pp. 1-21, Jan 2021.*

12. *P. Bedi and P. Gole. "Plant disease detection using hybrid model based on convolutional autoencoder and convolutional neural network." Artificial Intelligence in Agriculture, vol. 5, pp. 90-101, Jan 2021*

13. *Y. Li, J. Nie and X. Chao. "Do we really need deep CNN for plant diseases identification?" Computers and Electronics in Agriculture, vol. 178, pp. 105803-105818, Nov 2020*
.

14. *V. Prashanthi and K. Srinivas. "Plant disease detection using Convolutional neural networks." International Journal of Advanced Trends in Computer Science and Engineering, vol. 9, no. 3, pp. 2632-2637, May 2020.*

15. *P. Sharma, Y. P. Berwal and W. Ghai. "Performance analysis of deep learning CNN models for disease detection in plants using image segmentation." Information Processing in Agriculture, vol. 7, no. 4, pp. 566-574, Dec 2020.*

16. *S. Arivazhagan and S. V. Ligi "Mango leaf diseases identification using convolutional neural network." International Journal of Pure and Applied Mathematics, vol.120, no. 6, pp.1106711079, August 2018.*

17. *M. Agarwal, S. K. Gupta and K. K. Biswas. "Development of Efficient CNN model for Tomato crop disease identification." Sustainable Computing: Informatics and Systems, vol. 28, pp. 100407-100415, Dec 2020.*

18. *S. M. Hassan, M. Jasinski, Z. Leonowicz, E. Jasinska and A. K. Maji. "Plant disease identification using shallow convolutional neural network." Agronomy, vol. 11, no. 12, pp. 23882396, Nov 2021.*

19. *K. P. Ferentinos "Deep learning models for plant disease detection and diagnosis." Computers and electronics in agriculture, vol. 145, pp. 311-318, Feb 2018.*

20. *S. Kumar, V. Chaudhary, S. K. Chandra. "Plant disease detection using CNN.", Turkish Journal of Computer and Mathematics Education (TURCOMAT), vol. 12, no. 12, pp. 2106-2112, May 2021.*