International Journal of Advance and Innovative Research
Volume 11, Issue 1 (V): January - March, 2024
ISSN 2394 - 7780

A SURVEY ON CROPS DISEASE PREDICTION USING MACHINE LEARNING AND DEEP LEARNING

Dr. P. M. Hasabnis, Ms. Prerna S Dabhade, Ms. Pallavi G. Tayde and Ms. Megha R. Chopade
Department of Computer Science and Engineering, Mauli Group of Institution's, College of Engineering and Technology Shegaon, India

ABSTRACT

The crucial role of plants in climate, agriculture, and economies underscores the importance of their care. Similar to humans, plants are susceptible to diseases caused by bacteria, fungi, and viruses. Timely identification and treatment of these diseases are vital to prevent widespread destruction. This paper proposes a deep learning model for plant disease detection, aiming to accurately identify diseases in crops at an early stage. Early detection is essential for maintaining crop quality and yield by enabling appropriate treatments.However, disease detection requires specialized knowledge in plant pathology. The developed model utilizes neural networks, incorporating augmentation to expand the dataset. A Convolutional Neural Network (CNN) with multiple convolution and pooling layers is employed, trained on the PlantVillage dataset. Subsequently, the model undergoes rigorous testing, using 15% of the PlantVillage data, including images of healthy and diseased plants. The proposed model achieves a testing accuracy of 98.3%.In conclusion, this study focuses on a deep learning model for plant disease detection using leaves' images. Future integration with drones or other systems could enable real-time disease detection, reporting the location of diseased plants for prompt intervention.

Keywords: Deep Learning, Convolutional Neural Network, VGG, Resnet, PlantVillage, Crop disease.

I. INTRODUCTION

Crop identification is a crucial aspect of modern agriculture, optimizing resource allocation and estimating yields. With the global population on the rise, the demand for agricultural products is increasing rapidly. A vast amount of data is generated across various agricultural fields. Analyzing this data aids in predicting crop yield, assessing soil quality, anticipating plant diseases, and understanding how meteorological factors impact productivity. Effective crop protection is essential to sustain agricultural output. Pathogens, pests, weeds, and animals contribute to productivity losses. Crop diseases, stemming from pests, insects, and pathogens, can significantly reduce yields if not promptly addressed. Farmers incur financial losses due to these diseases. This paper presents a survey of diverse machine learning techniques employed for plant disease prediction. Automatic disease detection in plants facilitates early diagnosis and prevention, ultimately enhancing agricultural productivity.

II. LITERATURE REVIEW

A. Ip et al. (2018) conducted a comprehensive review on crop protection through the utilization of big data, emphasizing its role in weed control. The study delved into topics such as invasive species detection, forecasting and modelling herbicide resistance, support systems for crop protection, and robotic weed control. The paper also elucidated the machine learning methodologies employed to address these challenges.

B. Ebrahimi et al. (2017) introduced a method for detecting thrips (Thysanoptera) in crop canopies, specifically targeting identification within strawberry plants using Support Vector Machine (SVM) classification based on canopy images. The approach incorporates various kernel functions in SVM for parasite classification and Thysanoptera detection. Evaluation metrics including MAE, RMSE, MPE, and MSE were employed, revealing an error rate below 2.25% when utilizing color index and region index for classification. The removal of image backgrounds was achieved using MATLAB R2010a as part of the image processing technique.

C. Iqbal et al. (2018) focus on citrus plant diseases and their classification, detailing techniques for segmentation, feature extraction, feature selection, image processing, and classification. The research delves into automated tools for detection and classification, addressing diseases such as canker, black spot, citrus scab, melanose, and greening. In comparison with existing surveys, the K-mean algorithm is employed for disease extraction at different stages of analysis. For color feature computation and classification, Back Propagation Neural Network (BPNN) and Grey Level Co-Occurrences Matrix (GLCM) are utilized. The article covers techniques for preprocessing, including color-based transformation, image enhancement, noise reduction, resizing, and segmentation. Various feature extraction methods based on texture, color, and shape

ISSN 2394 - 7780

are discussed. The study also provides a summary of different classifier techniques and their applications, emphasizing that segmentation accuracy is enhanced through pre-processing techniques.

Algorithms. The integration of machine learning algorithms, including but not limited to support vector machines, decision trees, and deep learning models, allows for automated disease classification based on learned patterns from labelled datasets. Kulkarni et al.'s work likely builds upon existing literature, acknowledging the significance of datasets like Plant Village and leveraging techniques like transfer learning for enhanced model performance. As with many studies in this domain, challenges related to dataset diversity, model generalization to real-world conditions, and interpretability are prevalent. Future research directions may involve refining the synergy between image processing and machine learning, optimizing model architectures, and addressing the scalability of these approaches for large-scale agricultural applications. The study by Kulkarni et al. contributes to the growing body of literature, emphasizing the potential of image processing and machine learning for effective and automated plant disease detection in agricultural settings.

III. Image Analysis for Disease Detection in Plant

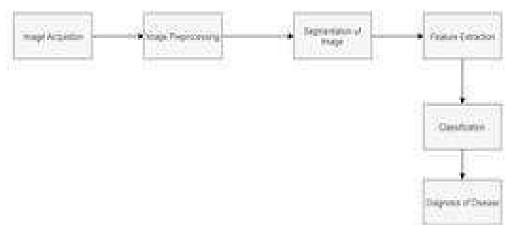

Fig. 1: General workflow of plant disease prediction model

The initial phase in image analysis is the image acquisition process, also referred to as digital image acquisition. This involves representing the visual character of an object through digital encoding, typically achieved by capturing an image using a camera. In contemporary times, digital imaging has extended to mobile phones, enhancing user-friendliness. The media utilized for image acquisition include photographs, printed paper, and photographic film, primarily capturing visual moments.In image preprocessing, two distinct types exist: digital image processing and analog image processing. The primary objective is the removal of unwanted features from the image, a process requiring various algorithms. The key steps in image preprocessing involve Image Acquisition, Image Normalization, Image Enhancement, Segmentation, and Morphology. Image segmentation, as highlighted by Oliver et al. (2018), entails separating an image into pixels and their similar attributes, aiding the image interpretation process. This transformation elevates the image from a low-level to a high-level representation, with the success of image analysis heavily reliant on the reliability of the segmentation process. Both contextual and non-contextual segmentation processes are employed, utilizing several algorithms.

Feature selection involves preserving a copy of the original features. In the subsequent feature extraction process, new feature sets are generated, focusing on eliminating unwanted noise and selecting necessary features for image analysis. This process includes the transformation of attributes, enhancing the speed and effectiveness of the overall procedure.The classification process categorizes data into multiple classes. In cases of new observations, determining their class assignment is crucial. Ferentinos is mentioned in relation to classification, but additional context is needed. Numerous classification algorithms are available, ensuring accurate classification results in this stage of image analysis.

Deficiencies. Each image is meticulously labelled with corresponding disease types, providing a comprehensive ground truth for the model. The dataset covers a wide range of crops, facilitating the creation of a robust and versatile predictive model capable of identifying and classifying different crop diseases accurately.by incorporating the PlantVillage dataset into our model development process, we ensure that the predictive capabilities of our model are well-informed and accurate, contributing to more effective crop management and disease mitigation strategies in agriculture.

IV. METHODOLOGY

A. CNN

The Convolutional Neural Network (CNN) is a specialized deep neural network designed for image recognition and classification. It processes data by scanning crop images from left to right and top to bottom, extracting pertinent features. These images, captured via cameras, drones, or other devices, serve as input. The CNN performs various operations through layers such as Convolutional, Pooling, and Fully Connected. The Convolutional layer generates an activation map by filtering the images pixel by pixel. The Pooling layer then reduces data size for more efficient storage. Lastly, the Fully Connected layer flattens the output from preceding layers into a single vector for the next stage.

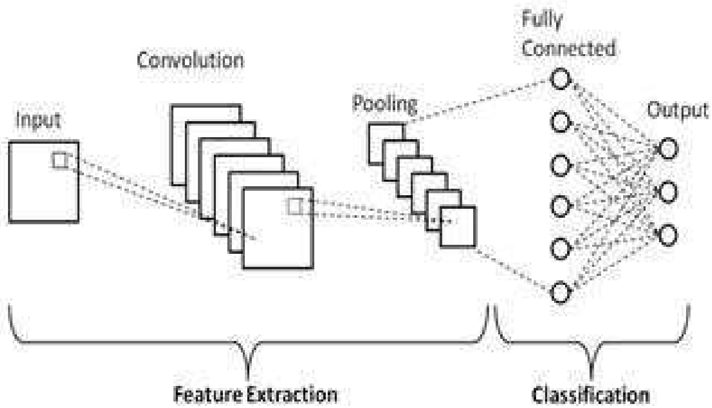

Fig. 2: Convolutional Neural Network

B. VGG

VGG, a deep convolutional neural network architecture, originated from the Visual Geometry Group at the University of Oxford. Recognized for its simplicity and effectiveness in image classification, it includes variants like VGG 16 and VGG 19. Although newer architectures like ResNet and Inception have surpassed VGG in efficiency, it remains a crucial reference in deep learning.

VGG 16 comprises 13 convolutional layers grouped into 5 blocks, each using 3x3 filters, a stride of 1, and padding of 1. Rectified Linear Unit (ReLU) activation functions follow each convolutional layer for non-linearity. After each block, a max-pooling layer with a 2x2 window and a stride of 2 reduces spatial dimensions. Fully connected layers also employ ReLU activation, and the final layer, with neurons equal to task classes, uses softmax to convert scores into class probabilities. The input image for VGG 16 is fixed at 224x224 pixels.

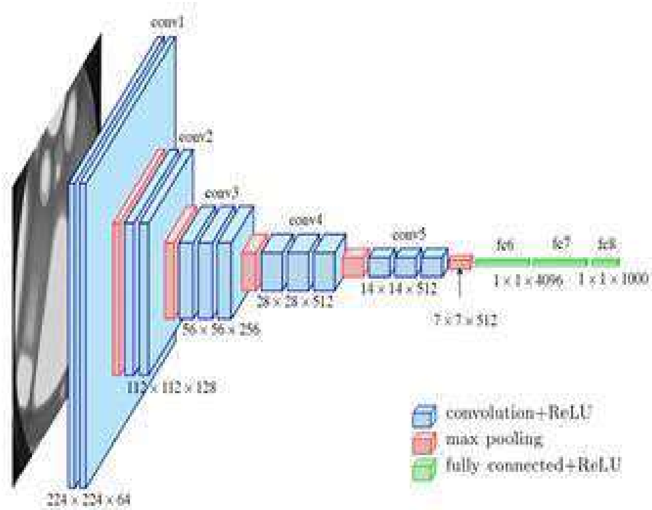

Fig. 3 VGG

C. RESNET

ResNet, short for "Residual Network," revolutionized computer vision and deep CNNs by addressing the challenge of vanishing/exploding gradients. It excels in tasks like image classification, object detection, and image segmentation. The key innovation is the introduction of Residual Blocks, utilizing skip connections in CNNs. Instead of a layer learning the complete mapping, ResNet allows the network to fit the residual mapping. This is expressed as \(H(x) = F(x) + x\), where \(F(x)\) represents the residual. Skip connections mitigate performance issues in individual layers, enabling successful training of very deep neural networks.

ISSN 2394 - 7780

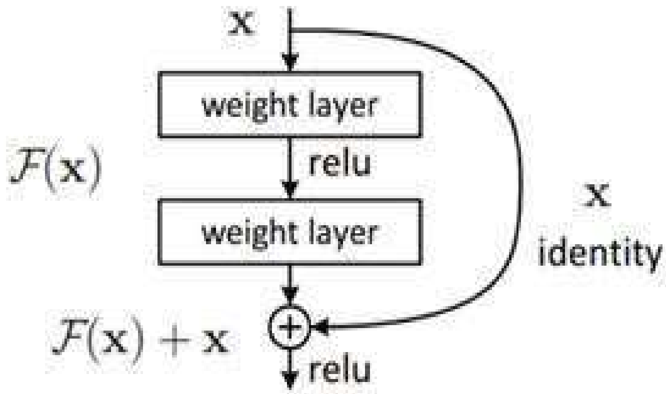

Fig. 4 RESNET

V. Dataset Discription

The investigation utilized plant leaf images sourced from the PlantVillage42 database, encompassing 8121 images of healthy bell pepper, potato, and tomato leaves. For disease detection, 31,061 images of diseased leaves were collected, covering bacterial spot in bell peppers, early blight caused by Alternaria tomatophila and Alternaria solani, late blight caused by Phytophthora infestans in potatoes or tomatoes, as well as bacterial spot and tomato mosaic virus in tomatoes (refer to Supplementary Table S7). In the case of tomatoes, four diseases were specifically chosen from a pool of nine to ensure accurate detection of common diseases across crops and identification of threats to farms.

Algorithm 1. Stepwise Disease Detection Model

Input: Healthy and diseased leaves of bell pepper, potato, and tomato

Output: Classified images as species and disease types of leaf

START:

1. Resize the image to 224 pixels
2. Split the data into 80% training set and 20% test set
3. 18 times data augmentation by rotating 20 degrees of training data

    #Step1 : Crop Classification

4. Start training of five pre-trained CNN models using 80% of training set
5. Perform validation and tune hyperparameter
6. Classification kinds of sample through an activation function, Softmax
7. Determine optimal model with high validation accuracy
8. Test with data not used to train the model
9. #Step2 : Disease Detection
10. Start training of five pre-trained CNN models using 80% of each crop's training set
11. Perform validation and tune hyperparameter
12. Determine sample as healthy or diseased through an activation function, Softmax
13. Determine optimal model with high validation accuracy
14. Test with data not used to train the model

    #Step3 : Disease Classification

15. Start training of five pre-trained CNN models using 80% of training set of each crop's disease images
16. Perform validation and tune hyperparameter
17. Classification kinds of disease through an activation function, Softmax
18. Determine optimal model with high validation accuracy
19. Test with data not used to train the model

END

Fig. 5 Stepwise disaease detection algorithm

The stepwise evaluation of the plant disease detection model involved using diseased image data from diverse crops to achieve a level of development suitable for smart farming. The evaluation encompassed apple, cherry, corn, grape, peach, and strawberry crops, ensuring a comprehensive assessment of the model's effectiveness across different plant types.

Table- I (a): Dataset Description

Class	Plant Name	Healthy or Diseased	Disease Name	Images (Number)
C_0	Apple	Diseased	Apple_scab	2016
C_1	Apple	Diseased	Black_rot	1987
C_2	Apple	Diseased	Cedar_apple_rust	1760
C_3	Apple	Healthy	-	2008
C_4	Blueberry	Diseased	-	1816
C_5	Cherry_(including_sour)	Diseased	Powdery_mildew	1683
C_6	Cherry_(including_sour)	Healthy	-	1826
C_7	Corn_(maize)	Diseased	Cercospora_leaf_spotGray_leaf_spot	1642
C_8	Corn_(maize)	Diseased	Common_rust	1907
C_9	Corn_(maize)	Diseased	Northern_Leaf_Blight	1908
C_10	Corn_(maize)	Healthy	-	1859
C_11	Grape	Diseased	Black_rot	1888
C_12	Grape	Diseased	Esca_(Black_Measles)	1920
C_13	Grape	Diseased	Leaf_blight (Isariopsis_Leaf_Spot)	1722
C_14	Grape	Healthy	-	1692
C_15	Orange	Diseased	Haunglongbing (Citrus_greening)	2010
C_16	Peach	Diseased	Bacterial_spot	1838
C_17	Peach	Healthy	-	1728
C_18	Pepper_bell	Diseased	Bacterial_spot	1913
C_19	Pepper_bell	Healthy	-	1988
C_20	Potato	Diseased	Early_blight	1939
C_21	Potato	Diseased	Late_blight	1939
C_22	Potato	Healthy	-	1824
C_23	Raspberry	Healthy	-	1781
C_24	Soybean	Healthy	-	2022
C_25	Squash	Diseased	Powdery_mildew	1736
C_26	Strawberry	Diseased	Leaf_scorch	1774
C_27	Strawberry	Healthy	-	1824
C_28	Tomato	Diseased	Bacterial_spot	1702

Utilizing this table provides insights into the quantity of images within each class, with approximately 2000 images per class. The dataset encompasses fourteen distinct plants, featuring both healthy and diseased leaf images for every plant. Predominantly, the dataset is rich in images of Tomato and Apple plants, while Raspberry,

Table- I (b): Dataset Description

Class	Plant Name	Healthy or Diseased	Disease Name	Images (Number)
C_29	Tomato	Diseased	Early_blight	1920
C_30	Tomato	Diseased	Late_blight	1851
C_31	Tomato	Diseased	Leaf_Mold	1882
C_32	Tomato	Diseased	Septoria_leaf_spot	1745
C_33	Tomato	Diseased	Spider_mites Two-spotted_spider_mite	1741
C_34	Tomato	Diseased	Target_Spot	1827
C_35	Tomato	Diseased	Tomato_Yellow_Leaf_Curl_Virus	1961
C_36	Tomato	Diseased	Tomato_mosaic_virus	1790
C_37	Tomato	Healthy	-	1926
Total				70295

Soybean, and Squash classes exhibit fewer images in comparison.

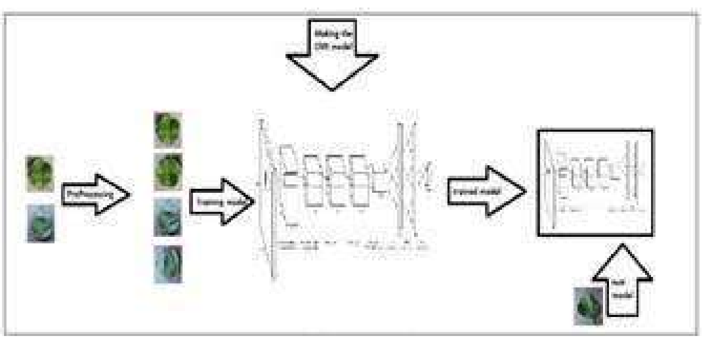

Fig. 6: Applied Methodologies

VI. DISCUSSION

This research underscores the significance of plant disease detection in contemporary times. The Deep Learning model, implemented in Python, underwent testing using 20% (14,059) images from the PlantVillage dataset, spanning 38 distinct classes. The test set comprised a random 20% selection from each class. Additionally, real-time images from the local environment, unrelated to the dataset's classes, were included. Despite challenges such as nighttime captures with flash and images with dirt, the model achieved over 95% accuracy, correctly classifying 96 out of 100 total images.

VII. CONCLUSION

This research employs deep learning techniques to establish an automated plant disease detection system. The system relies on a straightforward classification mechanism, leveraging CNN's feature extraction capabilities. For predictions, fully connected layers are employed. The study utilized a publicly available dataset comprising 70,295 images, with an additional 100 images from experimental conditions and the actual environment. The system demonstrated an impressive 98% testing accuracy on the publicly accessible dataset and performed well on images of plants from Sukkur IBA University. The conclusion drawn is that CNN is highly suitable for the automatic detection and diagnosis of plant diseases. The envisaged integration of this system into mini-drones for real-time disease detection in cultivated areas holds promise. Despite being trained on the Plant Village dataset with only 38 classes, the system can effectively identify whether a plant is diseased, as symptoms tend to be similar across different plant types. To enhance accuracy on real-condition images, future improvements could involve adding more actual environment images to the dataset, enabling the classification of additional plant and disease types. A proposed three-layer approach for the future involves the first layer detecting the presence of any plant in an image, the second layer determining the plant type, and the third layer identifying and classifying any diseases present.

REFERENCES

[1] Martinelli, F. et al. State-of-the-art approaches to plant disease detection: A comprehensive review. Agron. Sustain. Dev. 35, 1–25 (2015).

[2] Sankaran, S., Mishra, A., Ehsani, R. & Davis, C. Comprehensive review of advanced techniques for plant disease detection. Comput. Electron. Agric. 72, 1–13 (2010).

[3] Hasan, R. I., Yusuf, S. M. & Alzubaidi, L. Comprehensive analysis and discussion of the state-of-the-art in deep learning for plant diseases. Plants 9, 1302 (2020).

[4] Harakannanavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A. & Pramodhini, R. Detection of plant leaf diseases through the application of computer vision and machine learning algorithms. Glob. Trans. Proc. (2022).

[5] Yamashita, R., Nishio, M., Do, R. K. G. & Togashi, K. Convolutional neural networks: Overview and applications in radiology. Insights Imaging 9, 611–629 (2018).

[6] Zia, T. & Zahid, U. Architectures of long short-term memory recurrent neural networks for Urdu acoustic modeling. Int. J. Speech Technol. 22, 21–30 (2019).

[7] Dhaka, V. S. et al. Surveying deep convolutional neural networks applied to predict plant leaf diseases. Sensors 21, 4749 (2021).

[8] Lu, J., Tan, L. & Jiang, H. Review of convolutional neural network (CNN) applications in plant leaf disease classification. Agriculture 11, 707 (2021).

[9] Kurmi, Y. et al. Deep CNN model for detecting diseases in crops using leaf images. Multidimension. Syst. Signal Process. 33, 981–1000 (2022).

[10] Fang, Y., & Ramasamy, R. P. (2015). Current and prospective methods for detecting plant diseases. Biosensors, 5(3), 537-561.

[11] Kalpesh Shinde1, Nishant Dhamale2, Sudarshan Dangat3 , Prof. Anand Khatri4, "Crop Prediction and Plant Leaf Disease Prediction Using Deep Learning", 2022.

[12] Nikhil Patil, Rajab Ali. Vaibhav Wankhedkar. Prof. Deepali Nayak, "Crop Disease Detection using Deep Convolutional Neural Networks", 2019.