Green Guardian

Mrs. Prathibha T.S [prathibha024@gmail.com]

t Assistant Professor, Dept. of CSE, SIET

Mandara H [hmandhara$5@gmail.com], Priyanshu Prasad [8324priyanshu@gmail.com],

Shreyas Salunkhe [shreyassalunkhe2003@gmail.com], Rajan Kumar Sah [therajanshah@gmail.com

2 Department of Computer Science and Engineering

1-2 Shridevi Institute of Engineering and Technology

ABSTRACT

Plant diseases exert deep effects on the economy, impacting local and global scales. These diseases can lead to large losses within agricultural productivity, affecting crop yields and general quality. Deep learning algorithms, in this context, are widely known as effective solutions. However, the utilization of these particular black-box approaches raises several concerns regarding trust within interpreting and validating the decisions generated through the models. To classify and identify different ailments with better accuracy, this study proposes a plant disease classification system based on explainable artificial intelligence (XAI). The system identifies up to 38 different plant diseases with respective accuracy, precision, and recall as 99.69%, 98.27%, and 98.26%. These predictions are subjected for additional.

keywords—*Plant Disease Detection, Convolutional Neural Network, Prediction Model, Explainable AI, Deep Learning.*

I. INTRODUCTION

Farmers now face more difficulties than ever before due to the world's growing population and shifting climate. The detection and control of plant diseases is one of the major global challenges [1]. If left undiagnosed, these illnesses can cause significant drops in crop yield and quality, resulting in significant financial losses and endangering global food security [2]. Manual inspection by specialists is a common component of traditional disease diagnosis techniques, which is subjective and time-consuming [1].Furthermore, the disease may have spread widely by the time symptoms become apparent, which reduces the effectiveness of mitigation measures. Advances in technology have caused a major transformation in agriculture in recent years [3]. presenting ideas that have the potential to transform conventional farming methods.

II. LITERATURE REVIEW

Plant disease detection using machine learning has become a game-changing technology in agriculture, providing effective and precise ways to detect and treat crop diseases [4]. ML models can examine datasets to find subtle indications of diseases that the human eye might miss by utilizing sophisticated algorithms [8]. Here, we evaluate the performance of state-of-the-art algorithms for plant disease diagnosis. These algorithms were chosen due to their ease of implementation and the fact that most of them are open source. One subclass of deep neural networks created especially for image processing and recognition applications are convolutional neural networks (CNNs) [9]. CNNs are excellent at capturing hierarchical features and patterns because they are modeled after the visual processing found in the human brain. An input layer, several convolutional layers, pooling layers, and fully connected (dense) layers leading to the output layer make up a basic CNN model. The pooling layers down sample the spatial dimensions for more effective processing, while the convolutional layers apply filters to the input to extract features. Based on the features that were extracted, the dense layers subsequently carry out classification [10]. CNNS have demonstrated remarkable efficacy in computer vision applications, including facial recognition, object detection, and image classification. One of the strongest deep neural networks is the residual network (Rest Net) [11], which performs exceptionally well when used to solve classification problems. Rest Net, which is based on the CNN architecture, can support hundreds or even thousands of convolutional layers.

III. RELATED WORK

Detection of plant diseases is extremely crucial in agriculture. productivity significantly impacts the economy [28].Early detection and control of plant diseases prevent avoiding crop loss ensures food availability and sustaining farmers' livelihoods [22]. This section provides an overview of current ML model approaches for detecting plant disease. Ferentinos et al. [4] developed CNN models with variants, Google Net, OverFeat VGG for plant disease detection. Diagnosis from photos of healthy simple leaves, diseased plants. Training was on large database of 87,848 images of 25 plant species 58 [plant, disease] pairs. The top-performing model was 99.53% accurate in properly recognizing the combination or healthy plants. Mehedi et al. [14] introduced a transfer learning three pre-trained models (EfficientNetV2L, MobileNetV2 and ResNet152V2 identified 38 leaf diseases in 14 plants. The data were collected from Kaggle [15]. EfficientNetV2L did the best 99.63% accuracy. The XAI integration through LIM improves model clarity. Mohantyetal [16] employed a public dataset [17] containing 54,306 images of diseased and healthy leaves of plants. CNNs (Alex Net, Google Net) were trained to identify 14 crop species and 26 diseases with high accuracy.99.35% on the test set. Jasimetal. [18] investigated DL model applications in early disease detection and plant disease classification emphasizing the capability for greater precision than to classical ML.

Study	Dataset(s)	Methodology	Results	XAI Method	Mobile App	Classes	Species	Instances
Ferentinos et al. [4	Open Database	CNN and Variants	accuracy: 99.53%	No	No	58	25	[illegible]
Mehedi et al. [14]	Kaggle [15]	EfficientNetV2L, MobileNetV2, ResNet152V2	accuracy: 99.63%	Lime	No	38	14	54,305
Mohanty et al. [16]	PlantVillage [17]	AlexNet and GoogLeNet	accuracy: 99.35%	No	No	26	14	54,306
Jasim et al. [18]	PlantVillage [17]	CNN	accuracy: 98.029%	No	Yes	15	3	[illegible]
Ramesh et al. [19]	Random	Random Forest	accuracy: 70%	No	No	2	1	170
Harakannanavar et al. [20]	PlantVillage [17]	SVM, K-NN, CNN	accuracy: 99.6%	No	No	2	1	-
Fernandez et al. [21]	PlantVillage [17]	CNN	Lime is better	LIME, SHAP Grad-CAM	No	15	-	[illegible]
Kinger et al. [22]	PlantVillage [17]	VGG16	98.15%	Grad-CAM	No	38	14	>70,000
Nahiduzzaman et al. [23]	Real	Proposed CNN	accuracy: [illegible] ± 2.8%	SHAP	No	3	1	[illegible]
Arsenovic et al. [1]	PlantVillage [17]	Proposed: PlantDiseaseNet	accuracy: 93.67%	No	No	42	12	[illegible]
Khattak et al. [24]	Citrus [26], PlantVillage [17]	Proposed CNN	accuracy: 94.55%	No	No	5	1	2,293
Singh et al. [26]	Google Images + Ecosia [27]	MobileNet, Faster-RCNN	accuracy: increase by 31%	No	No	17	13	2,598

TABLE 1: RESEARCH MATRIX OF RELATED WORK

IV. PROPOSED METHODOLOGY

This section describes the proposed methodology in detail as shown in Fig.1.

A. DATA PRE-PROCESSING

In this step, background is removed from the images to separate the region of interest (ROI). The

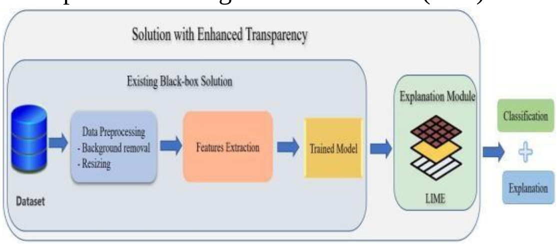

FIGURE 1: THE PROPOSED METHODOLOGY WORKFLOW

background can introduce noise into the data, leading to potential misdiagnoses. For instance, a leaf's shadow or the surrounding environment might be mistaken for a disease symptom, skewing the results. The images are also separated into distinct classes based on different crops to enhance model training efficiency. This step overall enhances the optimal accuracy and reliability of the trained classifier in disease detection, while decreasing the corresponding computation time. Some results before and after pre-processing are shown in Fig 2 and 3 respectively

B. FEATURES EXTRACTION

Step involves selecting key subset extracts features from available data for a ML model It is a critical step in the ML process since the A model's performance can be considerably affected by the Quality and quantity of features. Feature extraction by default method is transfer learning using the EfficientNetB0 model where the base model's layers are frozen to avoid training. The model's later improvement with global average pooling and a dense layer for multi-class classification is employed to create various features, improving the functionality of disease detection.

Plant	Disease	No of Samples	Plant	Disease	No of Samples
Apple	Apple Scab	2520	Potato	Early Blight	2424
Apple	Black Rot	2484	Potato	Late Blight	2424
Apple	Cedar Apple Rust	2200	Potato	Healthy	228[illegible]
Apple	Healthy	2510	Raspberry	Healthy	222[illegible]
Blueberry	Healthy	2270	Soybean	Healthy	2527
Cherry	Powdery Mildew	2104	Squash	Powdery Mildew	217[illegible]
Cherry	Healthy	2282	Strawberry	Leaf Scorch	221[illegible]
Corn (Maize)	Cercospora Leaf Spot / Gray Leaf Spot	2052	Strawberry	Healthy	2280
Corn (Maize	Common Rust	2384	Tomato		212[illegible]
Corn (Maize	Northern Leaf Blight	2385	Tomato		240[illegible]
Corn (Maize	Healthy	2324	Tomato		231[illegible]
Grape	Black Rot				
Grape	Esca (Black Measles	2400	Tomato	Septoria Leaf Spot	2181
Grape	Leaf Blight (Isariopsis Leaf Spot)	2152	Tomato	Spider Mites / Two-Spotted Spider Mite	217[illegible]
Grape	Healthy	2115	Tomato	Target Spot	228[illegible]
Orange	Haunglongbing (Citrus Greening)	2513	Tomato	Yellow Leaf Curl Virus	2451
Peach	Bacterial Spot	2297	Tomato	Mosaic Virus	223[illegible]
Peach	Healthy	2160	Tomato	Healthy	240[illegible]
Pepper, Bell	Bacterial Spot	2391			
Pepper, Bell	Healthy	2485			

TABLE 2: DATASET DESCRIPTION

FIGURE 2: RANDOMLY SELECTED PLANT IMAGES AFTER PREPROCESSING

FIGURE 3: RANDOMLY SELECTED PLANT IMAGES BEFORE PREPROCESSING

C. MODEL TRAINING AND VALIDATION

To ensure the robustness of our model, the dataset is split into an 80:20 ratio, with 80% of the images being used for training and the remaining 20% for validation.

V. RESULTS AND DISCUSSIONS

The methodology proposed employs four different state-of-the-art models viz., CNN, MobileNetV2, Efficient Net BO, and ResNet-50 for the detection of the diseases of the plants. The results in Table 4 indicate that Efficient Net BO is better than four other models with respect to accuracy, precision, and recall. MobileNetV2 is next after EfficientNetB0 with the best classification performance with 96.89%. ResNet-50 is doing worst compared to the rest with accuracy rate of 79.83%. These findings illustrate the efficiency of the EfficientNetB0 model with the highest accuracy amongst other models. The graphical view of the model accuracy is given in Fig. 4. There is evident variability in training and validation accuracy's values as the number of epochs increases. We have used XAI methods, specifically using the Lime framework, to increase the explainability of our machine learning models. This framework helped us describe our model's predictions, the useful insights into how the model made its predictions which depicts LIME explanations for predicted black-box model. These are Pepper, Bell with Bacterial Spot disease images with LIME explanations. In addition, 100% confidence indicates model predicted the disease with high probability.

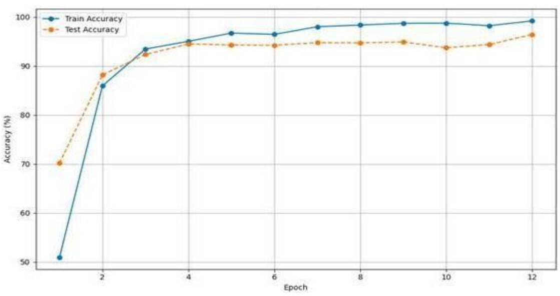

FIGURE 3:ACCURACY CURVE FOR PLANT DISEASE DETECTION

VI. STASTICAL ANALYSIS

Here, an analysis of variance (ANOVA) test [37]is performed to determine the statistical significance of the machine learning models. ANOVA is a statistical technique used to analyze differences between group means in a sample.

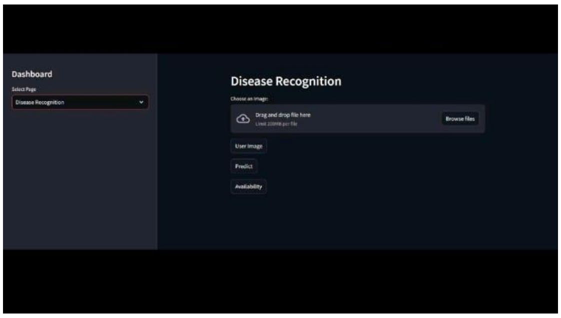

FIGURE 2:HOME PAGE OF DISEASE RECOGINIZATION

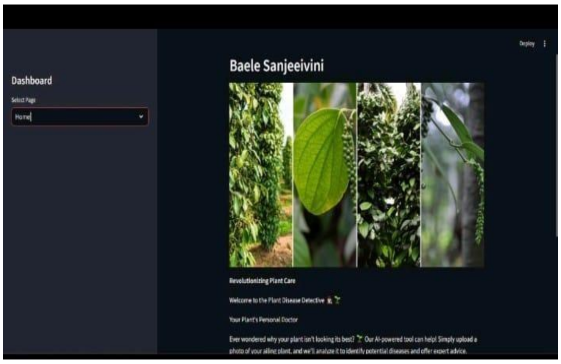

FIGURE 5: PLANT DISEASE DETECTION -USER INTERFACE

VII. VISUALIZATION

We also conducted an analysis with Power BI to obtain insights from our dataset, representing them through different types of charts and graphs. Number of samples per plant are represented in Fig. 8, while accuracy obtained by each plant are represented in Fig. 9. These values can also relies on the quality of the study has some limitations. First of all, we did not considered large or varied datasets. Secondly, our method used four pre-trained models for comparison. Expanding the model to incorporate more sophisticated pre-trained models can improve classification performance. Finally, using a greater number of training data can possibly produce better results.

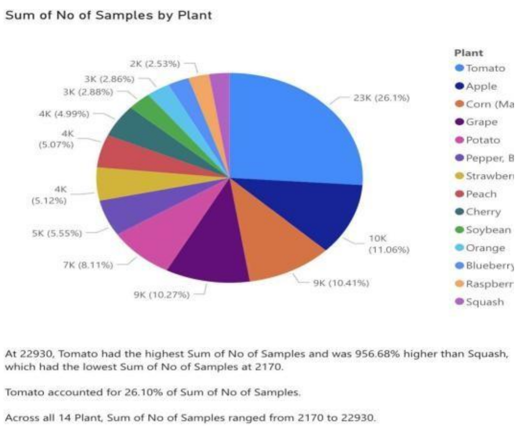

FIGURE 6: NO. OF SAMPLES PER PLANT

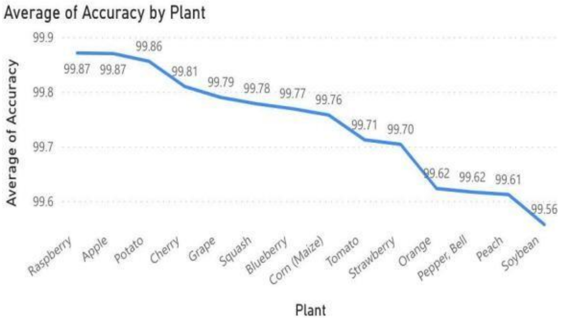

FIGURE 7: ACHIEVED ACCURACY FOR EACH PLANT

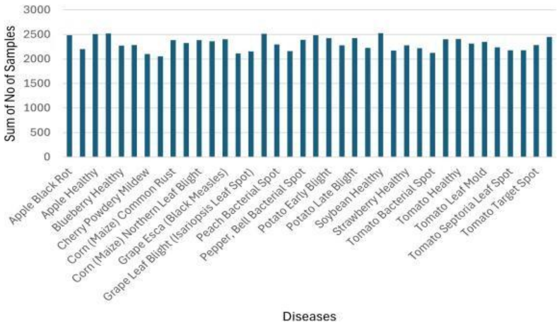

FIGURE 8: NO. OF SAMPLES BY DISEASES

VIII. THREATS TO VALIDITY

The main purpose of this work is to help farmers early diagnose plant diseases. However, there are some limitations in this research. First, we did not consider large or varied datasets. Secondly, our method was based on four pre-trained models for comparison. Using the model to incorporate more sophisticated pre-trained models might improve classification ability. Finally, using a greater amount of training data might lead to better results.

External validity is a major concern because most datasets are collected in controlled settings with clear, isolated leaves. However, real-world conditions—like leaves overlapping, dirt, insect damage, or varied weather lighting—are much messier. If the model is not exposed to such diversity during training, its accuracy in real farming environments could drop significantly. Additionally, new plant species or emerging diseases not present in the training data could go undetected

IX. USER INTERFACE

The User Interface is designed to be highly intuitive, clean, and accessible for a wide range of users — from farmers with basic tech skills to professional agronomists and researchers.

A. Upload Interface

The interface's main purpose is to enable users to submit infected plant pictures for examination. This module's salient features include.

1) **Drag-and-Drop File Upload**: Users can upload image files instantly by simply dragging and dropping them into a designated area. This streamlines interaction and does away with the need to browse through file explorers.

2) **Browse/Choose File Button:** manually choose an infected image from local storage, users who prefer more conventional approaches can also use the "Choose File" button.

3) **Input Validation:** Only image files in supported formats (such as.jpg,.jpeg, and.png) are accepted, according to input validation. Without requiring a server round trip, unsupported formats immediately produce feedback messages [10].

X. LIMITATIONS & CHALLENGES

1. Limited Diversity of the Dataset

Most datasets for plant diseases are captured in controlled environments, with clean, centered images. In field situations, leaf images can be blurry, occluded, captured in low-light conditions, or contain complex backgrounds. This difference degrades model accuracy when put to use in the field.

2. Generalization to Novel Diseases

The CNN model can only identify diseases it has been trained on. When a new disease or a variation of an existing disease occurs, the model can refuse to identify it properly or incorrectly identify it as another disease.

3. Overfitting Risk

Since CNNs are very resourceful, they tend to overfit the training data, particularly when the data set is minimal or homogenous. This results in good training performance but low real-world performance.

4. Limitations of Explainability

Although methods such as LIME offer explainability, they are far from ideal. Occasionally, the highlighted regions are not entirely significant to human consumers, resulting in suspicions or misinterpretation regarding the model's decision-making process.

5. High Computational Resources

Training deep CNN models is computationally intensive and necessitates strong hardware such as GPUs. Even at inference time, execution of large models on low-end phones or embedded systems is not feasible without model compression.

6. Environmental Sensitivity

Environmental conditions like noise in the background (soil, sky, multiple leaves), seasonal variation, and varying plant stages can mislead the model, reducing its dependability under diverse real-world scenarios.

7. Data Labelling Errors

Incorrectly identified training images will cause significant model performance damage. High-quality expert-validated data is time consuming and costly.

8. Restrictive Multimodal Analysis

The proposed CNN-based method analyze only the visual features. Plant health is also a function of non-visual factors such as temperature, soil moisture, and nutrient level, which the pure image-based methods overlook.

9. Misinterpretations from User Interface

Users may misinterpret the model's confidence scores or visual explanations if the UI does not make them evidently clear, resulting in incorrect decisions even when correct predictions are made.

X.CONCLUSION

Plant diseases are a major threat to our economy, resulting in huge losses in agricultural output and affecting the livelihood of farmers. Plant disease control is not only an agricultural issue but also a strategic step towards economic growth. In this context, this study introduces the efficacy of a deep learning-based plant disease detection system using explainable artificial intelligence (XAI). The use of sophisticated deep learning models not only improves the accuracy of disease identification but also offers interpretability through the use of XAI technique .EfficientNetB0 is used in this paper to train the ML model using a dataset of 87,000 images. The trained model shows competence in correct classification of 38 different types of diseases with accuracy, precision, and recall rates of 99.69%, 98.27%, and 98.26%, respectively. In addition, the LIME framework is used to offer meaningful explanations that facilitates informed decision-making. The visual explanations not only exhibit the effective generalization of the model but also expose biases learned from outlier images. These findings enable researchers and field practitioners to better understand the reasoning behind the classification of plant diseases, illuminating the inner workings of the black box model. Besides statistic alanalyses, particularly ANOVA, demonstrates models' performance of significant models. In the future, there is potential for the development of a more robust model that considers diseases affecting
numerous plant varieties. In addition, the aggregation of written reports presenting disease findings in technical and non-technical terms is another task that will assist the adoption of the model. Additionally, the inclusion of Internet of Things (IoT) devices will be responsible for the complete automation of the disease detection systems on farms.

This paper offers a reliable and interpretable plant disease recognition system based on deep learning, that is, a Convolutional Neural Network (CNN) integrated with the Local Interpretable Model-agnostic Explanations (LIME) approach. The system is highly accurate, precise, and recalls with confidence, proving to have a high potential for real-world deployment in agriculture. The system offers not only predictions but also visual explanations, thereby increasing user trust and enabling informed decision-making by farmers and agricultural experts.

In spite of its success, issues like dataset limitations, environmental variations, and requirements for better interpretability still persist. Future research will be aimed at increasing the dataset, integrating multimodal data, optimizing the model for mobile deployment, and improving explanation methods. In all, this research is a promising step toward intelligent, reliable, and accessible plant disease management systems that can significantly contribute to global food security and sustainable agriculture.

ACKNOWLWDGEMENT

We also acknowledge the publicly available datasets and open-source communities whose contributions greatly facilitated the development and evaluation of the plant disease detection system. Finally, we extend our appreciation to our peers, colleagues, and agricultural experts whose insights and encouragement were instrumental in refining this study.

REFERENCES

[1] M. Arsenovic, M. Karanovic, S. Sladojevic, A. Anderla, and D. Stefanovic, "Solving current limitations of deep learning based approaches for plant disease detection," Symmetry, vol. 11, no. 7, p. 939. Jul. 2019.

[2] M. Nagaraju and P. Chawla, "Systematic review of deep learning techniques in plant disease detection," Int. J. Syst, Assurance Eng. Manag.,vol. 11, pp. 547-560, May 2020.

[3] L. Li, S. Zhang, and B. Wang, "Plant disease detection and classification by deep leurning-A review," IEEE Access, vol. 9, pp. 56683-56698, 2021.

[4] K. P. Ferentinos, "Deep leaming models for plant disease detection and diagnosis," Comput. Electron. Agricult., vol. 145, pp. 311-318,Feb. 2018.

[5] M. H. Saleem, J. Potgieter, and K. M. Arif, "Plant disease detection and classification by deep learning," Plants, vol. 8, no. 11, p. 468, 2019.

[6] N. Nigar, M. Umar, M. K. Shahzad, S. Islam, and D. Abalo, "A deep learning approach based on explainable artificial intelligence for skin lesion classification," IEEE Access, vol. 10, pp. 113715-113725,2022.

[7] M. Tan and Q, Le,"EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int. Conf. Mach. Learn., 2019,pp. 6105-6114.

[8] N. Nigar, A. Jaleel, S. Islam, M. K. Shahzad, and E. A. Affum, "IoMT meets machine learning: From edge to cloud chronic diseases diagnosis system," J. Healtheare Eng., vol. 2023, no. 1, 2023, Art. no. 9995292.

[9] K. O'Shea and R. Nash, "An introduction to convolutional neural networks," 2015, arXiv:1511.08458.

[10] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," IEEE Trans, Neural Netw. Learn. Syst., vol. 33, no. 12, pp. 6999-7019, Dee. 2022.

[11] K. He, X. Zhang, S. Ren, and J, Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR),Jun. 2016, pp. 770-778.

[12] D. Sinha and M. El-Sharkawy, "Thin Mobile Net: An enhanced MobileNet architecture," in Proc. IEEE 10th Annu. Ubiquitous Comput., Electron.Mobile Commun. Conf. (UEMCON), Oct. 2019, pp. 0280-0285.

[13] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why should i trust you?'Explaining the predictions of any classifier," in Proc. 22nd ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining, 2016, pp. 1135-1144.

[14] M. H. K. Mehedi, A. K. M. S. Hosain, S. Ahmed, S. T. Promita,R. K. Muna, M. Hasan, and M. T. Reza, "Plant leaf disease detection using transfer learning and explainable AI," in Proc. IEEE 13th Annu.Inf. Technol., Electron. Mobile Commun. Conf. (IEMCON), Oct. 2022,pp. 0166-0170.

[15] S. Bhattarai. New Plant Diseases Dataset. [Online]. Available:https://www.kaggle.com/datasets/vipooool/new-plant-diseases-dataset

[16] S. P. Mohanty, D. P. Hughes, and M. Salathé, "Using deep learning for image-based plant disease detection," Frontiers Plant Sci., vol. 7, p. 1419, Sep. 2016.

[17] S. Mohanty, https://github.com/salathegroup/plantvillage_deeplearning_paper_dataset

[18] M. A. Jasim and J. M. Al-Tuwaijari, "Plant leaf diseases detection and classification using image processing and deep learning techniques,"in Proc. Int. Conf. Computer Sci. Softw. Eng. (CSASE), Apr. 2020, pp. 259-265.

[19] S. Ramesh, R. Hebbar, M. Niveditha, R. Pooja, N. Shashank, and P. V. Vinod, "Plant disease detection using machine learning," in Proc.Int. Conf. Design Innov. 3Cs Compute Communicate Control (ICDI3C), Apr. 2018, pp. 41-45.

[20] S. S. Harakannanavar, J. M. Rudagi, V. I. Puranikmath, A. Siddiqua, and R. Pramodhini, "Plant leaf disease detection using computer vision and machine learning algorithms," Global Transitions Proc., vol. 3, no. 1,pp. 305-310, Jun. 2022.