



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** X **Month of publication:** October 2024

DOI: <https://doi.org/10.22214/ijraset.2024.64917>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Plant Diseases Prediction Using Machine Learning

Urvashi Waghmare¹, Samiksha Sorte², Adyashri Shinde³, Dhanashree Nikhate⁴, Prof. Mahesh Dumbere⁵

^{1, 2, 3, 4}Students, ⁵Assistant professor, Department of Computer Science and Engineering, Rajiv Gandhi College of Engineering, Research and Technology, Chandrapur, Maharashtra, India

Abstract: *The early detection and accurate prediction of plant diseases are crucial for improving crop health and maximizing agricultural productivity. Traditional methods of disease detection, which rely heavily on manual observation, are often time-consuming, labor-intensive, and prone to human error. Recent advancements in machine learning (ML) have opened new possibilities for developing efficient, automated systems that can predict plant diseases with high accuracy. This paper explores various machine learning techniques, including supervised and deep learning models, for plant disease prediction. By analyzing features such as leaf texture, color, and environmental data, these models can identify patterns indicative of specific diseases. Several datasets, including real-time image data and environmental metrics (such as humidity, temperature, and soil quality), are utilized to train and evaluate the models. The study focuses on key ML algorithms like Support Vector Machines (SVM), Decision Trees, Random Forests, Convolutional Neural Networks (CNN), and Transfer Learning to understand their efficacy in disease prediction. These models are evaluated based on their accuracy, precision, recall, and computational efficiency. The results demonstrate that deep learning models, particularly CNNs, outperform traditional methods in image-based plant disease identification. In addition, environmental data integration improves the predictive power of ML models, enabling proactive disease management strategies.*

Keywords: *Machine Learning, Plant Disease Prediction, Convolutional Neural Networks, Supervised Learning, Precision Agriculture, Deep Learning.*

I. INTRODUCTION

Agriculture plays a vital role in sustaining the global economy and food supply, but plant diseases pose a significant threat to crop production, often resulting in severe yield losses and economic setbacks. The ability to detect and manage plant diseases in their early stages is critical for ensuring food security and maintaining crop health. Traditional methods of plant disease detection rely heavily on visual inspection by experts, which can be time-consuming, subjective, and prone to human error. Moreover, these methods are often not feasible for large-scale agricultural operations, where timely intervention is essential for preventing widespread damage.

In recent years, technological advancements, particularly in the fields of artificial intelligence (AI) and machine learning (ML), have shown immense promise in addressing these challenges. Machine learning algorithms can analyze complex patterns in large datasets and make accurate predictions, offering an automated, scalable solution to plant disease detection. By using image recognition, sensor data, and environmental inputs, ML models can identify symptoms of plant diseases at an early stage, enabling proactive management strategies and reducing the need for excessive pesticide use.

Several types of machine learning techniques, ranging from traditional methods like Decision Trees and Support Vector Machines (SVMs) to advanced deep learning architectures such as Convolutional Neural Networks (CNNs), have been employed to predict plant diseases. CNNs, in particular, have demonstrated remarkable success in image classification tasks and are widely used for analyzing leaf images to detect symptoms like discoloration, spots, or lesions that indicate the presence of diseases.

This paper aims to explore the application of machine learning models in plant disease prediction, focusing on their potential to enhance agricultural practices. We investigate a variety of algorithms and their performance in detecting and predicting plant diseases using both image-based data and environmental factors such as humidity, temperature, and soil quality. Furthermore, we discuss the challenges faced in deploying ML models in agricultural settings, including the need for large labeled datasets, the risk of overfitting, and the integration of these systems with real-time monitoring technologies such as the Internet of Things (IoT).

The objective of this research is to evaluate the effectiveness of ML-based systems in predicting plant diseases, highlight their advantages over traditional methods, and propose directions for future work that can further enhance the precision and applicability of these technologies in modern agriculture. Through this study, we aim to contribute to the growing body of knowledge on precision agriculture and support the development of intelligent farming systems that can mitigate the impact of plant diseases on global food production.

II. RELATED WORK

The literature on leaf disease detection using deep learning is reviewed in this part, along with other essential research. This section reviews the use of deep learning for detecting leaf diseases and highlights key research in the field. In recent years, artificial intelligence (AI) has been widely applied in healthcare and agriculture. AI has been especially useful in analyzing images, both in medical applications and in farming.

For example, Sanjiv Sannakki and his team used AI and image processing to help diagnose plant diseases. Similarly, Monika Jhuria and her colleagues used image processing to grade fruits and identify diseases. They used an artificial neural network to help classify the different diseases. Kaiyi Wang and his team developed a new approach using image processing and computer vision to identify insect pests and plant diseases. They studied images of vegetable plants taken by smartphones to assess the health of the plants.

Researchers have explored various methods and models for detecting plant diseases using machine learning. One study focused on distinguishing healthy leaves from diseased ones using image processing and machine learning techniques. Many diseases cause leaves to lose chlorophyll, leading to dark or black spots, which can be detected through machine learning. These techniques include classification, feature extraction, image preprocessing, and image segmentation. For feature extraction, the Grey Level Co-occurrence Matrix (GLCM) is often used. The Support Vector Machine (SVM) is a popular machine learning method for classification, while the Convolutional Neural Network (CNN) technique has further improved disease recognition accuracy.

the accuracy and performance of different methods for detecting plant diseases using machine learning and deep learning techniques.

For apple leaves, using a Convolutional Neural Network (CNN) resulted in an impressive 99% overall accuracy. The classification of plants was found to be 97.71% accurate. Another study focused on detecting rice leaf diseases using machine learning. They identified three common rice diseases: leaf blight, bacterial leaf blight, and brown spot. Clear images of damaged rice leaves were used, and after pre-processing, the dataset was trained using several machine learning techniques, such as K-Nearest Neighbour (KNN), Decision Tree (J48), Bayesian Network, and Logistic Regression. After testing, the decision tree method had over 97% accuracy.

Wheat leaf diseases were also studied, as these diseases can harm wheat production. Although algorithms can detect common wheat leaf diseases, the challenges lie in the wide range of possible infections, such as viruses, bacteria, fungi, and insects. Identifying these diseases is important for farmers to monitor large wheat fields. The study used machine learning to classify and detect wheat leaf diseases, addressing the main challenges farmers face.

In another study, a CNN architecture was suggested for detecting plant leaf diseases, achieving a classification accuracy of 95.81%. For maize plants, different machine learning methods were tested using photos of the plants. The Random Forest (RF) algorithm performed best, with an accuracy of 79.23%. These models help farmers identify and classify diseases early, preventing the spread of new infections.

Image processing techniques are also used for disease detection. This method includes several steps: first, preprocessing the image, then using K-means clustering to segment the leaf and find damaged areas. After that, features like texture data are extracted, and finally, the disease is classified using Support Vector Machine (SVM).

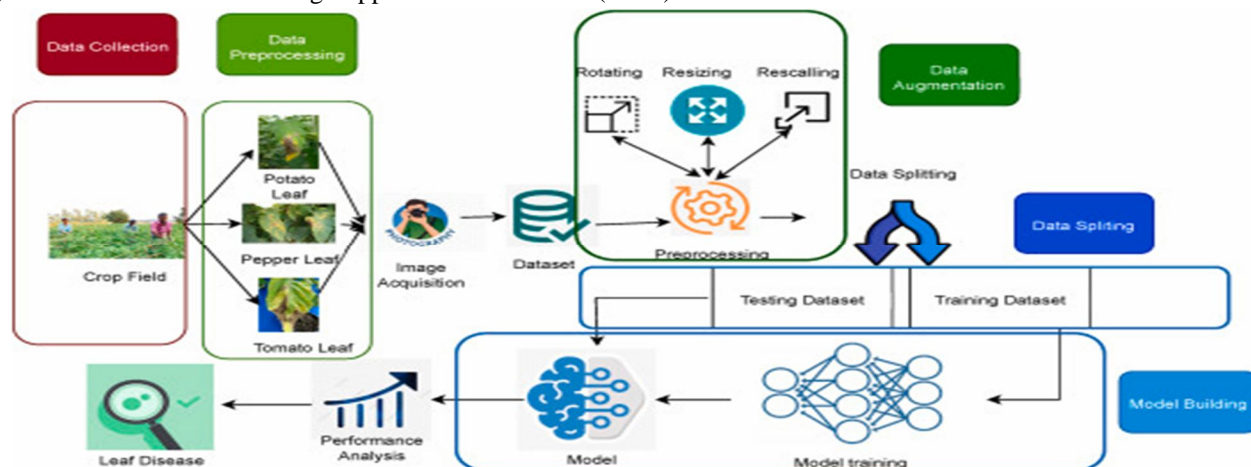


Fig: 1 work flow diagram

III. METHODOLOGY

Dataset: The project begins by gathering images of different types of leaves, such as potato, tomato, and pepper leaves. These images can be captured using a camera or a smartphone. For training the deep learning model, a publicly available dataset of leaf images was used. This dataset was used during both the testing and training stages of the deep learning framework to help the model learn to recognize and classify diseases in the leaves.

Processing: Preprocessing is an important step to clean up and prepare the raw images before using them in the deep learning model. The images collected may contain "noise," like extra details, poor lighting, or unwanted elements that could affect the model's ability to learn accurately.

- 1) *Rotation:* Some images might be captured at different angles. By rotating the images, the model can learn to recognize the leaf no matter how it is positioned. This makes the model more flexible in understanding different orientations.
- 2) *Resizing:* The images may have different sizes, but for the model to work properly, all images need to be the same size. Resizing adjusts the images to a uniform size, ensuring they fit the model's requirements.
- 3) *Shearing:* Shearing slightly shifts parts of the image, changing the shape without distorting the important features. This helps the model learn to recognize leaves even if they appear in slightly altered shapes.

A. Training and Building the Model

Training Phase: In this phase, the Transfer Learning (TL) models are trained using a set of images from the training dataset. These images are used to "teach" the model how to recognize patterns, such as the different types of leaves and diseases. The model learns by adjusting its internal settings based on the training data.

Validation Phase: After the training is done, the model is tested using a separate set of images, called the test dataset. These images were not used during training, so they help evaluate how well the model performs on new, unseen data. This phase checks the accuracy and effectiveness of the model by measuring how well it identifies leaf diseases based on the test images.

B. Model Construction

- 1) Collecting images from the dataset
- 2) Pre-process image data by resizing and rotating images.
- 3) Creating convolute feature connect into Fully Connected Layers. Usually, it is flattened, converted to a one-dimensional (1D) array (or vector), and then joined to one or more completely connected layers.
- 4) Finally, extract the features for different classes of the input.

C. Model Evaluation

- 1) From an ideal dataset, 80% of photos are taken for training and 20% for testing.
- 2) Validation data is used to check accuracy by applying the predict function and accurately extracting features.
- 3) Images are taken for confirming detection once validation provides good results.
- 4) Finally, characteristics are retrieved to determine whether or not the leaves are infected.

D. Performance Evaluation

- 1) **Accuracy:** This tells us how many predictions the model got correct out of all the cases.
- 2) **Precision:** This measures how often the model was correct when it predicted a leaf had a disease.
- 3) **Recall:** This shows how good the model is at finding all the diseased leaves in the dataset.
- 4) **F1 Score:** This combines precision and recall into a single score to give a balanced view of the model's performance.
- 5) **Training Accuracy:** This measures how well the model performed on the training data (the data it learned from).
- 6) **Training Loss:** This shows how much error the model made while learning from the training data. Lower loss is better.
- 7) **Validation Accuracy:** This measures how well the model performs on new, unseen data (the test set) during evaluation.
- 8) **8 Validation Loss:** This is the error the model made on the test data. Again, lower loss means the model is performing better.

E. Algorithms

- 1) **Support Vector Machine (SVM):** Classifies diseases by finding the best decision boundary.
- 2) **K-Nearest Neighbours (KNN):** Classifies based on the closest examples in the dataset.
- 3) **Decision Trees (DT):** Splits data into branches to classify diseases.

- 4) Random Forest (RF): Combines multiple decision trees for more accurate classification.
- 5) Convolutional Neural Networks (CNN): Deep learning algorithm that processes images for high-accuracy disease detection.
- 6) Artificial Neural Networks (ANN): Learns patterns in data, though less effective for images than CNN.
- 7) Naive Bayes: Classifies based on probabilities of features, useful for simple tasks.
- 8) Logistic Regression: Predicts disease based on the probability of input features.
- 9) Transfer Learning (AlexNet, VGG, ResNet): Pre-trained CNN models fine-tuned for plant disease detection.
- 10) K-Means Clustering: Segments images by clustering pixels, often used in preprocessing.

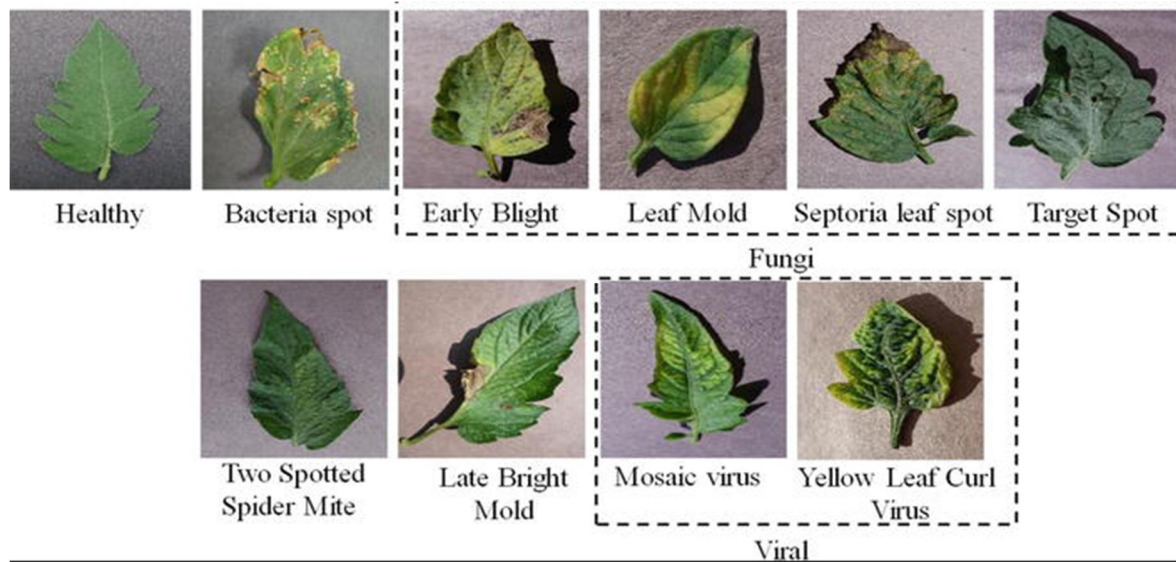


Fig .2 Example of diseased image

IV. RESULT AND ANALYSIS

The average training and validation accuracy and loss for the three experiments. For any deep learning algorithm, learning curves represent the learning capabilities during the training concerning the dataset in an incremental fashion. With the increasing number of epochs, the training accuracy accurately interprets how well the model is learning with the training dataset. The validation accuracy, on the other hand, provides a prediction of the model's generalizability based on a hold-out validation dataset. explain it in simple. n deep learning, learning curves (like accuracy and loss graphs) show how well a model learns from data over time as it trains.

Training Accuracy: This measures how well the model is doing with the data it has already seen during training. As the model goes through more training cycles (epochs), the training accuracy should generally increase, showing that the model is learning the patterns in the training data.

Validation Accuracy: This checks the model's performance on new, unseen data (the validation set). It helps us see if the model can generalize what it's learned to data it hasn't seen before. If the validation accuracy is high, the model is likely making good predictions on new data.

Loss: This is a measure of how well (or poorly) the model is doing at each stage. Lower loss means better performance.

It represents the training and validation accuracy and loss respectively of ResNet50. The loss curve represents that the validation loss and training loss decreased over time and the time between them is minimal during the experiments though we found little fluctuation during the validation test. Fig. 4 shows the training and validation accuracy and loss for the proposal. The validation and training accuracy curves in 4 depict that the accuracy performance is increasing with the training time although the validation curve shows some fluctuation. The loss graph on the other hand represents good fitting as both the training and validation of loss curves are suited well with almost no difference between them.

It represents the results considering accuracy, precision, . It is also clear that for any model ResNet50 outperforms the other model. To validate by ResNet50 more properly, we applied 5-fold cross validation approach on each of the when we deal with the ResNet50 model. Therefore, in that case, 80% data are used for training cases and the rest of 20% for testing purposes. It depicts the performance results (accuracy) on each of the folds. By averaging the accuracies of each fold we got average accuracy of 93.13%, 96.126% and 98.288% .

We focus on three major crops in this study: tomatoes, potatoes, and peppers the three most important crops in Bangladesh. Fig. 6 shows the summary of different leaves considering the accuracy performance of each of the models during the 3 experiments. From the graphs, it is clear that the detection accuracy of ResNet50 is higher than any models the average accuracy of the models. According to our findings the average accuracy for any kind of crop leaves, average accuracy for is better. This is due to the fact that the accuracy performance rises in a linear fashion with the size of the training set. It represents the confusion matrix for the CNN model for the testing case.

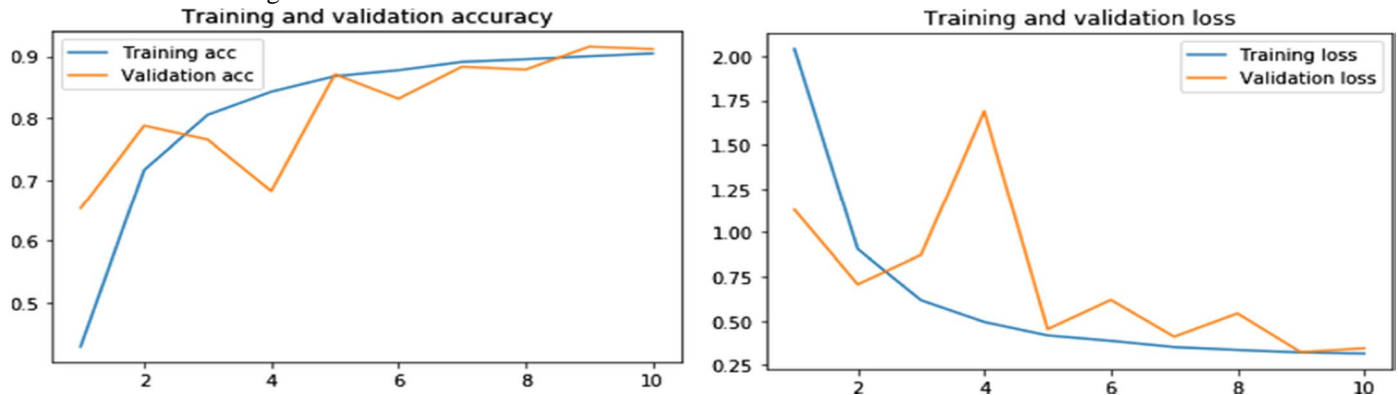


Fig . 3 Graph of training and validation

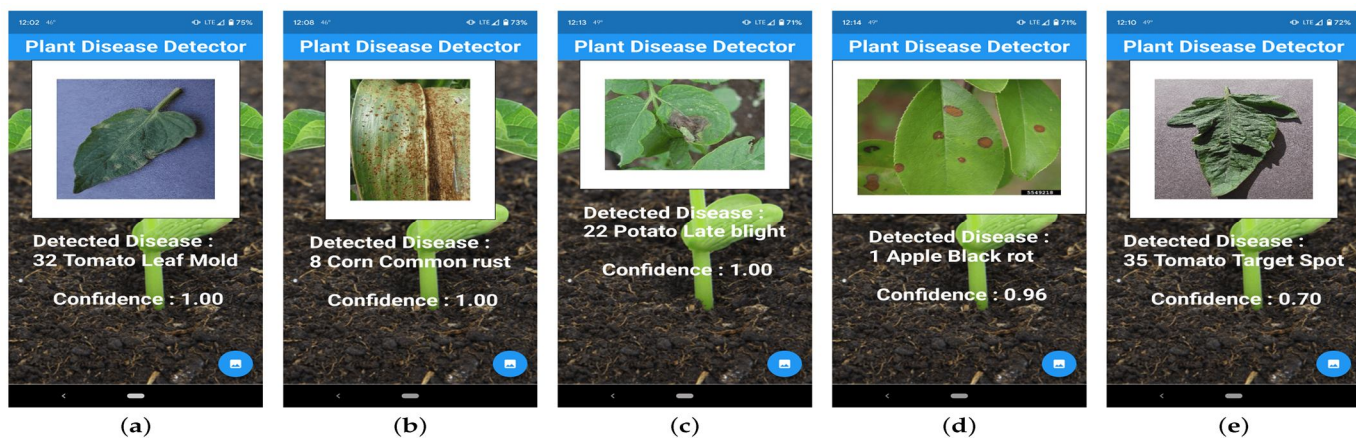


Fig . 4 Detected diseases

V. CONCLUSION

Plant disease diagnosis is an important process to determine the standard of the crop through which many factors such as yield capacity, richness of the grains, nutrition retention are evaluated. This paper aims to make a comprehensive study on various computational methods embedded in the plant disease identification and classification system. Many intelligent algorithms played a vital role in achieving the desired task. In the best performing deep learning and machine learning algorithms are plotted to portray the performance of each algorithm. Also, many other fusion models were built to improve the predictability of the computational model and are discussed in detail. The outcome of this study reveals the significance of automated tools for assisting the end users to find the plant disease without the human intervention. In future, prescriptive models are to be developed, which is in great demand near in future. The key objective of this work is to analyze different machine learning techniques widely used in the prediction of plant diseases and how advancement can be made in the future in this technique to achieve higher accuracy, robustness, cost-efficient disease prediction system. The steps involved in image processing techniques like pre-processing, segmentation, extracting feature and classification based on symptoms in the plant are discussed in this survey. Machine learning techniques play a key role in the machine vision system. In the future, deep learning framework can be used for disease prediction system. Integrating image processing techniques and deep learning techniques proved to be more potential in disease prediction system. Still, more investigations have to be made in these techniques for achieving better prediction system.



REFERENCES

- [1] Faisal Ahmed, Hawlader Abdullah Al-Mamun, ASM Hossain Bari, Emam Hossain, and Paul Kwan. Classification of crops and weeds from digital images: A support vector machine approach. *Crop Protection*, 40:98–104, 2012.
- [2] P.W. Bidwell, The agricultural revolution in new england, *Am. Hist. Rev.* 26 (4) (1921) 683–702.
- [3] Z.B. Husin, A.Y.B.M. Shakaff, A.H.B.A. Aziz, R.B.S.M. Farook, Feasibility study on plant chili disease detection using image processing techniques, in: 2012 Third International Conference on Intelligent Systems Modelling and Simulation, IEEE, 2012, pp. 291–296.
- [4] Hassan Asadollahi, Morteza Sabery Kamarposhty, and Mir Majid Teymoori. Classification and evaluation of tomato images using several classifier. In 2009 International Association of Computer Science and Information Technology-Spring Conference, pages 471–474. IEEE, 2009.
- [5] Davoud Ashourloo, Hossein Aghighi, Ali Akbar Matkan, Mohammad Reza Mobasheri, and Amir Moeini Rad. An investigation into machine learning regression techniques for the leaf rust disease detection using hyperspectral measurement. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 9(9):4344–4351, 2016.
- [6] J M.A. Talukder, M.M. Islam, M.A. Uddin, A. Akhter, M.A.J. Pramanik, S. Aryal, M.A. A. Almoyad, K.F. Hasan, M.A. Moni, An Efficient Deep Learning Model to Categorize Brain Tumor Using Reconstruction and Fine-Tuning, *Expert Systems with Applications*, 2023, 120534.
- [7] J S. Pawar, S. Shedje, N. Panigrahi, A. Jyoti, P. Thorave, S. Sayyad, Leaf disease detection of multiple plants using deep learning, in: 2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON), IEEE, 2022, pp. 241–245, 1.
- [8] J A. Dixit, S. Nema, Wheat leaf disease detection using machine learning method-a review, *Int. J. Comput. Sci. Mobile Comput.* 7 (5) (2018) 124–129.
- [9] Meng-han Hu, Qing-li Dong, Bao-lin Liu, and Pradeep K Malakar. The potential of double k-means clustering for banana image segmentation. *Journal of Food Process Engineering*, 37(1):10–18, 2014.
- [10] DS Guru, PB Mallikarjuna, S Manjunath, and MM Shenoi. Machine vision based classification of tobacco leaves for automatic harvesting. *Intelligent Automation & Soft Computing*, 18(5):581–590, 2012.
- [11] J K. Simonyan, A. Zisserman, Very Deep Convolutional Networks for Large-Scale Image Recognition, 2014 arXiv preprint arXiv:1409.1556.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)