IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10597

PLANT DISEASE DETECTION

Rohan Singhal[a], Shrey Bishnoi[b], Pankaj Gupta[c], Dr. Jyoti Kaushik[1]

[a,b]CSE Department, Maharaja Agrasen Institute Of Technology, India

[1]Assistant Professor, CSE Department, Maharaja Agrasen Institute of Technology, New Delhi, India

Abstract- Plant diseases threaten agricultural productivity and food security. Early detection and management of leaf diseases are crucial for minimizing crop losses. We propose an automated machine learning and image processing approach for leaf disease detection. We collected a comprehensive dataset of plant leaf images and pre-processed it to enhance image quality and extract relevant features. We then trained and classified the leaf images into different disease categories using machine-learning algorithms, including CNNs and SVMs...

To improve the detection accuracy, we incorporated image-processing techniques such as segmentation, feature extraction, and color analysis. These techniques enabled the extraction of disease-specific characteristics from the leaf images; further enhancing the classification performance. The proposed system achieved promising results, with an overall accuracy of 99.9% in detecting and classifying plant leaf diseases. We also developed a user-friendly interface that allows farmers to upload leaf images and receive instant disease diagnosis and recommended treatment strategies.

The potential impact of this research is significant, as it equips farmers with a reliable tool for early disease detection, enabling proactive measures to prevent the spread of diseases and optimize crop management practices. By minimizing crop losses and improving disease management, this approach contributes to sustainable agriculture and food production. Future work involves expanding the dataset, exploring more advanced machine learning algorithms, and integrating remote sensing technologies for real-time disease monitoring.

Keywords: plant leaf diseases, disease detection, machine learning, image processing, convolutional neural networks, support vector machines, crop management, sustainable agriculture, food security.

I.INTRODUCTION:

The primary source of food, revenue, and employment is agriculture, which makes up a substantial portion of the global economy. As in other low -and middle - income nations with large farmer populations, agriculture accounts for 18 % of India 's GDP increases earnings and raises employment to 53%.The gross value added (GVA) by agriculture to the national economy during the last three years has grown from 17.6% to 20.2%. The greatest portion of economic growth is provided by this industry. Therefore, the impact of plant disease and pest infections on agriculture may have an influence on the global economy by lowering the quality of food produced. Epidemic and endemic disease cannot be prevented using prophylactic medicines.

Each infection and pest condition, in particular, leaves behind distinct patterns that may be utilised to identify anomalies. Expertise and personnel are needed in order to identify a plant disease. Additionally, manual examination is subjective and time-consuming when determining the kind of plant infection, and occasionally the illness recognised by farmers or specialists may be deceptive. This might result in the use of an inappropriate medicine during the evaluation of the plant disease, which could lower the quality of the crops and ultimately pollute the environment.

Since the infected spots are originally seen as spots and patterns on leaves, there are several solutions to address the detection challenges for plants thanks to the development of computer vision.

Researchers to precisely identify and categorise plant diseases have put several methods forth. Some employ conventional image processing methods that include manual, or handcrafted, feature extraction and segmentation. A K-means clustering approach was presented by **Dubey et al. [1]** to partition the diseased region of leaves, with the multi-class

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10597

support vector machine (SVM) that was used to classify the data at the end. Probabilistic neural networks were utilised by **Yun et al. [2]** to extract statistical and meteorological characteristics.

The development of machine learning and deep learning has led to significant advances in plant disease detection. These approaches can automatically extract features from images, making it easier to identify and classify diseases. Additionally, deep learning models can be trained on large datasets, which can improve their accuracy. However, more research is needed to develop more diverse datasets, as current datasets may not be representative of the real world.

II.LITERARY SURVEY:

P. Krithika et al. [3] pre-processed images by resizing, enhancing contrast, and converting color space. They used K-Means clustering for segmentation and GLCM for feature extraction. Classification was done using multiclass SVM. **R. Meena et al. [4]** converted leaf colors into LAB* space after color space conversion and enhancement. They employed K-Means clustering for segmentation and utilized GLCM and SVM for feature extraction and classification, respectively.

Bharat et al. [5] captured images using a digital camera and enhanced them using a median filter. They performed segmentation with K-Means clustering and employed SVM for classification. **Pooja et al. [6]** conducted segmentation to identify infected regions of interest. They used K-Means clustering, Otsu's detection, and converted RGB to HSI for segmentation, utilizing boundary and spot detection algorithms.

Rukaiyya et al. [7] performed pre-processing through contrast adjustment and normalization. They transformed color into YCBCR and used bi-level thresholding. GLCM and HMM were used for feature extraction and classification, respectively. **Chaitali et al. [8]** applied image segmentation for background subtraction. Classification was done using KNN, ANN, and SVM methods. KNN classifies samples based on the nearest distance between trained and testing subjects.

Anjali et al. [9] developed a model using thresholding technique and morphological operations for extraction. They employed multiclass SVM as the classifier. Segmentation was based on color and luminosity components analysis in LAB* color spaces. GLCM was used for feature extraction. **Vijai Singh et al. [10]** captured samples of plant leaves (rose/beans, lemon, banana, beans) using a digital camera. They used a thresholding algorithm to identify green regions as background. Segmentation was performed using a genetic algorithm. Color co-occurrence was used for feature extraction. The Minimum Distance Criterion and SVM classifier were employed for classification, achieving an average accuracy of 97.6%.

Sa'ed Abed et al. [11] improved the quality of input samples through scaling and stretching. They created an HIS model and performed segmentation using combined Euclidean distance and K-means clustering. GLCM was used for feature extraction, and SVM for classification. **Arya et al. [12]** transformed RGB images and converted them to HIS format. They segmented components using Otsu's method.

Nema et al. [13] analyzed a database of 81 images in Lab color space. They performed segmentation using k-means clustering and disease classification using SVM. Statistical information such as mean, median, mode, and standard deviation were utilized. **Vidyashree Kanbur et al. [14]** developed a model for leaf disease detection using multiple descriptors. The model was tested on a local leaf database and showed superior performance. It can be further tested on publicly available datasets. **Pushpa et al. [15]** used Indices Based Histogram technique for segmenting unhealthy regions of leaves. Their segmentation technique outperformed other methods such as slice segmentation, polygon approximation, and mean-shift segmentation.

Kaleem et al. [16] pre-processed images by resizing, noise removal, brightness enhancement, and contrast adjustment. They employed K-means clustering for segmentation, utilized Statistical GLCM for feature extraction, and used SVM for classification of leaf disorders.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10597

Image Classification performed by **Sunil S. Harakannanavar et al. [17]** on the Plant Village dataset was assessed using sensitivity. The experiment was performed using SVM, KNN, and CNN. It was observed that the CNN (soft) classifier achieved the best accuracy of 99% for classifying the Plant leaf disease from all the experiments conducted above. CNN model got the best results out of all the experiments conducted. It was followed by SVM and KNN in order of ranking as observed from the experiments conducted.

Table 1: literary survey

Year	Title	Author	Description
2017	Leaf disease detection on cucumber leaves using multiclass support vector machine	P. Krithika et al. [3]	They used K-Means clustering for segmentation and GLCM for feature extraction. Classification was done using multiclass SVM.
2018	Detection of unhealthy plant leaves Using image processing and genetic algorithm with Arduino	Anjali et al. [9]	They employed multiclass SVM as the classifier. Segmentation was based on color and luminosity components analysis in LAB* color spaces. GLCM was used for feature extraction
2020	Detection of Leaf Disease Using Hybrid Feature Extraction Techniques and CNN Classifier	Vidyashree Kanbur et al.[14]	Developed a model for leaf disease detection using multiple descriptors. The model was tested on a local leaf database and showed superior performance. It can be further tested on publicly available datasets
2021	A Modern Approach for Detection of Leaf Diseases Using Image Processing and ML Based SVM Classifier	Kaleem et al [16]	Pre-processed images by resizing, noise removal, brightness enhancement, and contrast adjustment. They employed K-means clustering for segmentation, utilized Statistical GLCM for feature extraction, and used SVM for classification of leaf disorders.
2022	Plant leaf disease detection using computer vision and machine learning algorithms	Sunil S. Harakannanavar et al.[17]	The proposed model (DWT+PCA+GLCM+CNN) using computer vision and ML classification technique is used.

From the above research papers, we inferred:

- During the process of the project, data pre-processing is crucial as it increases the data efficiency and decreases response time.
- To provide the best result more models should be compared.
- We selected the plant village data set, which was used previously by **Sunil S. Harakannanavar et al [17]**. in 2022 for better comparison and survey.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10597

III.DATA PRE-PROCESSING:

About Dataset

The dataset used for this project has been taken from Plant-Village- Dataset that can be found on kaggle. The data used for this project is extracted from the folder named "color" which is situated in the folder named "raw" in the Repository. The Data fed for the modelling is of pepper, tomato, potato, strawberry, grape, corn, and apple Leaves. For training purpose, the Dataset comprises of 2 folders named Diseased and Healthy which contains images of leaves with respective labels. The Diseased Folder contains diseased/unhealthy, affected by Scab, Black Rot or Rust. The Healthy Folder consists of Green and healthy images.

Steps involved:

a) Loading Original Image. A total of 800 images for each class Diseased and Healthy is fed for the machine.

b) Conversion of image from RGB to BGR. Since Open CV (python library for Image Processing), accepts images in RGB colouring format so it needs to be converted to the original format that is BGR format.

c) Conversion of image from BGR to HSV. The simple answer is that unlike RGB, HSV separates luma, or the image intensity, from Chroma or the colour information. This is very useful in many applications. For example, if you want to do histogram equalization of a color image, you probably want to do that only on the intensity component, and leave the color components alone. Otherwise, you will get very strange colors. In computer vision, you often want to separate color components from intensity for various reasons, such as robustness to lighting changes, or removing shadows. Note, however, that HSV is one of many color spaces that separate color from intensity .HSV is often used simply because the code for converting between RGB and HSV is widely available and can be easily implemented.

d) Image Segmentation for extraction of Colors. In order to separate the picture of leaf from the background segmentation has to performed, the color of the leaf is extracted from the image.

e) Applying Global Feature Descriptor. Global features are extracted from the image using three feature descriptors namely :

• Color: Color Channel Statistics (Mean, Standard Deviation) and Color Histogram

• Shape: Hu Moments, Zernike Moments

• Texture: Haralick Texture, Local Binary Patterns (LBP)

IV.FEATURE EXTRACTION:

After extracting the feature of images the features are stacked together using numpy function "np.stack". According to the images situated in the folder the labels are encoded in numeric format for better understanding of the machine.
The Dataset is splitted into training and testing set with the ratio of 80/20 respectively.

(I) Feature Scaling

Feature Scaling is a technique to standardize the independent features present in the data in a fixed range. It is performed during the data pre-processing to handle highly varying magnitudes or values or units. If feature scaling is not done, then a machine-learning algorithm tends to weigh greater values, higher and consider smaller values as the lower values, regardless of the unit of the values.

Here, we have used Min-Max Scaler. This scaling brings the value between 0 and 1.

(II) Saving the Features

After features are extracted from the images, they are saved in HDF5 file. The Hierarchical Data Format version 5

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10597

(HDF5), is an open source file format that supports large, complex, heterogeneous data. HDF5 uses a "file directory" like structure that allows you to organize data within the file in many different structured ways, as you might do with files on your computer.

V.ALGORITHMS OVERVIEW:

(I) Logistic Regression (LR)

Logistic Regression is a machine-learning algorithm that can be applied to image classification tasks. In the context of image classification, Logistic Regression estimates the probability of an image belonging to a specific class based on its features. These features are typically extracted from the image using techniques like pixel intensity values or deep learning features. Logistic Regression learns the relationship between the extracted features and the probability of the image belonging to a particular class. It is a simpler and interpretable method compared to more complex algorithms like neural networks, making it suitable for smaller image classification tasks or when interpretability is important.

(II) Convolutional neural networks (CNN)

A **neural network** in which at least one layer is a **convolutional layer**. A typical convolutional neural network consists of some combination of the following layers:

- **convolutional layers**
- **pooling layers**
- **dense layers**

Convolutional neural networks have had great success in certain kinds of problems, such as image recognition.

(III) Linear Discriminant Analysis (LDA)

Linear Discriminant Analysis (LDA) is a statistical technique commonly used for image classification. It aims to find a lower-dimensional space that maximizes the separation between different image classes. LDA calculates class means and scatter matrices to capture the variance and covariance information within and between classes. It then projects the image features onto discriminant axes to transform the data. A classification rule can be applied to assign class labels to new images. LDA reduces dimensionality and improves classification accuracy by maximizing class reparability.

(IV) Random Forest (RF)

Random Forest is a popular ensemble learning method for image classification. It combines multiple decision trees to make predictions and handles complex image data effectively. By using random subsets of training data and image features, it reduces overfitting and improves generalization. Random Forests can handle high-dimensional feature spaces, capture feature interactions, and provide feature importance insights. They are robust against noise and outliers and have fast training and prediction times. Random Forests have been successfully used in various image classification tasks, making them a powerful and widely applied technique.

(V) Naïve Bayes

Naïve Bayes is a simple and efficient algorithm for image classification. It assumes feature independence and calculates probabilities based on Bayes' theorem. It estimates the probability of an image belonging to a class given its features. Naïve Bayes is computationally lightweight and suitable for large datasets. However, it may not capture complex feature interactions. It serves as a baseline method and performs well for simpler image classification tasks.

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10597

(VI) Decision Trees

Decision Trees are widely used for image classification due to their ability to handle complex feature interactions and produce interpretable classification rules. They recursively partition the feature space based on discriminative features, creating homogeneous subgroups within each partition. By following decision rules at each node, an image is classified based on the leaf node it reaches. Decision Trees capture non-linear relationships and handle high-dimensional feature spaces. However, they can over fit and techniques like pruning, ensemble methods, and regularization are used to improve performance. Overall, Decision Trees offer interpretability and effective classification for image data.

(VII) K nearest Neighbors (KNN)

K Nearest Neighbors (KNN) is a straightforward algorithm for image classification. It assigns a class label to an image based on the labels of its nearest neighbors in the feature space. The algorithm calculates distances between the image's feature vector and those of the training images. The k nearest neighbors are chosen, and the class label is determined by majority voting. KNN is simple to understand and implement, but the choice of k and feature selection greatly influence its performance. Additionally, it can be computationally demanding for large datasets.

(VIII) Support vector machines (SVM)

Support vector machines (SVMs) are a type of supervised machine learning algorithm that can be used for both classification and regression problems. However, they are most commonly used for classification problems. In image classification, SVMs can be used to identify objects or scenes in images. SVMs are a powerful tool for image classification, but they can be computationally expensive to train. Additionally, they can be sensitive to the choice of features. However, with careful training, SVMs can achieve high accuracy in image classification tasks.

VI.EVALUATION METRICS

F1 Score: Compute the F1 score, also known as balanced F-score or F-measure.

The F1 score can be interpreted as a harmonic mean of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. The formula for the F1 score is

F1 = 2 * (precision * recall) / (precision + recall)

In the multi-class and multi-label case, this is the average of the F1 score of each class with weighting depending on the average parameter.

VII.RESULT:

The result of proposed model is compared with existing models. It is observed that, the accuracy of proposed model (CNN) provides better accuracy of 99.93% compared to the other existing models.

Techniques	Accuracy
Naive Bayes	85.55%
Linear Discriminant Analysis	90.16%
Classification And Regression Tree	91.56%
Logistic Regression	91.72%
K-Nearest Neighbor	92.11%
Support Vector Machine	92.27%
Random Forest	96.41%
Convolutional Nueral Network	**99.93%**

Figure 1 : Comparison of different classification techniques

IARJSET ISSN (O) 2393-8021, ISSN (P) 2394-1588

International Advanced Research Journal in Science, Engineering and Technology

ISO 3297:2007 Certified ⁜ Impact Factor 8.066 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 10, Issue 5, May 2023

DOI: 10.17148/IARJSET.2023.10597

The proposed model (CNN) using computer vision and ML classification technique is compared with the methodology explained by **Hossain et al.[18]**, and **Vidyashree et al.[14]**, and **Thanjai Vadivel et al.[19]**, and **Sunil S. Harakannanavar et al.[17]**, and is tabulated in Figure 2. The accuracy obtained by the proposed method is better compared with the existing methodologies.

Authors	Techniques	Accuracy
Sunil S. Harakannanavar et al.,	CNN	99.09%
Hossain et al.,	SVM	90.00%
Vidyashreeet al.,	SVM	90.00%
Thanjai Vadivel et al.,	Fast Enhanced Learning Method	99.00%
Proposed Technique	**CNN**	**99.93%**

Figure 2 : Comparison of existing methodologies with proposed model

The above results were obtained on a system with the following specification:

- Intel ® core(™) i5-1035G1 CPU @ 3.37 GHz processor
- 8 GB SODIMM RAM
- Intel ® UHD Graphics
- NVIDIA GeForce MX230

VIII.CONCLUSION AND FUTURE SCOPE

The task of identifying leaf diseases requires loading the original images, converting them from RGB to BGR to HSV, performing image segmentation to determine the colour of the leaf, applying global feature descriptors to extract features from the image, encoding the labels in numeric format, dividing the dataset into a training set and a testing set, performing feature scaling on the training set, saving the features to an HDF5 file, and training eight different models.

The analysis of the proposed model is well suited for CNN machine learning classification technique with a desired accuracy compared to other state of the art method. In future, the model can be improved using fusion techniques for extraction of significant features and examined for other leaf samples of datasets.

IX.REFERENCES:

[1] Dubey, A. Kumar and Shanmugasundaram M. "Agricultural Plant Disease Detection and Identification," AgriSciRN: Plant Pathology (Sub-Topic) (2020): n. pag..

[2] Yun, "Detection and measurement of paddy leaf disease symptoms using image processing," 2017 *International Conference on Computer Communication and Informatics (ICCCI)*, pp. 1-4, 2017.

[3] P. Krithika and S. Veni, "Leaf disease detection on cucumber leaves using multiclass Support Vector Machine," 2017 *International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET)*, Chennai, India, 2017, pp. 1276-1281, doi: 10.1109/WiSPNET.2017.8299969.

[4] R. Meena , "modern approach for plant leaf disease classification which depends on leaf image processing," *IEEE International Conference on Computer Communication and Informatics* (2017), pp. 12-16

[5] N. Amoda, B. Jadhav, P. Kurle, S. Kunder, S. Naikwadi, 2014, "DETECTION AND CLASSIFICATION OF PLANT DISEASES BY IMAGE PROCESSING," *International Journal Of Engineering Research & Technology* (IJERT) ICONET – 2014 (Volume 2 – Issue 04),

[6] Pooja, V. et al. "Identification of plant leaf diseases using image processing techniques." 2017 *IEEE Technological Innovations in ICT for Agriculture and Rural Development (TIAR)* (2017): 130-133.

[7] Dhole, Sampada & Shaikh, Rukaiyya. (2016). "Review of Leaf Unhealthy Region Detection Using Image Processing Techniques," Bulletin of Electrical Engineering and Informatics. 5. 10.11591/eei.v5i4.498.

DOI: 10.17148/IARJSET.2023.10597

[8] Dhaware, Chaitali G., and K. H. Wanjale. "A modern approach for plant leaf disease classification which depends on leaf image processing." 2017 *International Conference on Computer Communication and Informatics (ICCCI)*. IEEE, 2017.

[9] Arya, M. S., K. Anjali, and D. Unni. "Detection of unhealthy plant leaves using image processing and genetic algorithm with Arduino." *2018 International Conference on Power, Signals, Control and Computation (EPSCICON*). IEEE, 2018.

[10] V. Singh, and A. K. Misra. "Detection of plant leaf diseases using image segmentation and soft computing techniques." Information processing in Agriculture 4.1 (2017): 41-49.

[11] S. Abed, A Esmaeel "A novel approach to classify and detect bean diseases based on image processing," *IEEE Symposium on Computer Applications & Industrial Electronics* (2018), pp. 297-302.

[12] M Arya, K Anjali, D Unni ,"Detection of unhealthy plant leaves Using image processing and genetic algorithm with *Arduino*," *IEEE International Conference on Power, Signals, Control and Computation* (2018), pp. 1-5.

[13] S. Nema, A. Dixit, "Wheat Leaf Detection and Prevention Using Support Vector Machine," *International Conference on Circuits and Systems in Digital Enterprise Technology* (2018), pp. 1-5 .

[14] V. Kanabur, Sunil S. Harakannanavar, Veena I. Puranikmath, Dattaprasad ,"Torse Detection of Leaf Disease Using Hybrid Feature Extraction Techniques and CNN," *Classifier Springer Comput*. (2019), pp. 1213-1220.

[15] Pushpa B. R., Shree Hari AV, and Adarsh Ashok. "Diseased leaf segmentation from complex background using indices based histogram." 2021 *6th International Conference on Communication and Electronics Systems (ICCES*). IEEE, 2021.

[16] M. K. Kaleem, N. Purohit, K. Azezew, S. Asemie ,"A modern approach for detection of leaf diseases using image processing and ML based SVM classifier," Turkish J. Comput. Math. Educ., 12 (13) (2021), pp. 3340-3347.

[17] Harakannanavar, S. S., Rudagi, J. M., Puranikmath, V. I., Siddiqua, A., and Pramodhini, R.,"Plant leaf disease detection using computer vision and machine learning algorithms," Global Transitions Proceedings, vol. 3, no. 1, pp. 305–310, 2022. doi:10.1016/j.gltp.2022.03.016.

[18] S. Hossain, R. Mou, M. Hasan, S. Chakraborty, "A Razzak Recognition and detection of tea leaf's diseases using support vector machine," *IEEE Int. Colloquium Signal Process. Appl.* (2018), pp. 150-154.

[19] Thanjai Vadivel, R. Suguna ,"Automatic recognition of tomato leaf disease using fast enhanced learning with image processing," Taylor Francis, Acta Agricult. Scandinavica, Sect. B — Soil Plant Sci., 71 (1) (2021), pp. 1-13.

