



e-ISSN: 2319-8753 | p-ISSN: 2320-6710

# IJRSET

International Journal of Innovative Research in  
**SCIENCE | ENGINEERING | TECHNOLOGY**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH


IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 9, Issue 11, November 2020

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA

**Impact Factor: 7.512**

 9940 572 462

 6381 907 438

 [ijirset@gmail.com](mailto:ijirset@gmail.com)

 [www.ijirset.com](http://www.ijirset.com)

# Plant Disease Detection and Classification using Random Forest Classifier Model.

Tanvi Talekar<sup>1</sup>

IT Student, Department of Information Technology, B.K. Birla College of Arts, Science and Commerce(Autonomous), Kalyan, Maharashtra, India<sup>1</sup>

**ABSTRACT:** Agriculture is the main stay of India's economy. Detection of plant disease at its early stage is important to maintain its quality and productivity and it also decrease the agricultural loss. Detection of plant disease is done using various automation techniques that mechanically finds and classify the disease and it is very beneficial as it reduces large work of detecting large farms and gives accurate results. It involves various stages like Preprocessing of image based dataset, Feature Extraction, training the classifier and classification. the dataset of healthy and unhealthy leaves are trained under random forest classifier for classification of healthy and unhealthy leaves also Random forest classifier provide more accuracy. Basically machine learning gives a proper way to identify the disease present in the plants.

**KEYWORDS:** Random Forest Classifier, Feature extraction, PreProcessing, Training, Classification, Plant disease detection.

## I. INTRODUCTION

Agriculture is a backbone of our country. Economy of India is highly dependent on agriculture. Therefore it is very important to maintain the quality of plant so for this detection of disease plays an important role. It becomes difficult for farmer to maintain each and every plant and detect every disease and it will consume more time and will not give accurate result so to avoid this automatic disease detection technique is used. Disease can be detected at early stage with the help of automatic disease detection technique. Booming cultivation of crops the recognition of plants, leaves and detection of the disease, symptoms of disease plays an important role. To avoid human interference automatic disease detection technique is used to detect and classify the disease and provide accurate result.

Various modern approaches came into existence for identification of plant disease. Various researches have done under the field of machine learning for plant disease detection using various machine learning approaches such as Random forest, K-Nearest Neighbour, Decision Tree, Navie Bayes, SVM etc.

Random forest is used for classification, regression and other task. Random forest handles both numeric and categorical data. We have applied Global feature descriptors. Global features are extracted from the images using three descriptors:

1. Hu Moments
2. Haralick Texture
3. Color Histogram

Hu moments is basically used to describe the outline of a leaf. Haralick texture is used to distinguish between the texture of healthy and unhealthy leaf. Color histogram used to represent the color of image.

The following paper is organized as follows: section II consists of related Plant Disease Detection studies. sections III, IV include the methodology and experimental result of the proposed study respectively, and finally section V concludes the study along with future scope.

## II. RELATED WORK

In[1]Devraj et.al. detected Various illness such as Alternaria Alternata , Bacterial Blight and Cercospora Leaf spot using image processing in MATLAB. So first leaf picture information assortment is done then pre-processing of the picture then the segmentation of the picture using K-means cluster methodology then GLCM is used for extracting the feature and at last the random forest classifier used for Coaching and testing the leaf of plant.In[3] Sardeshmukh et.al. developed a web application to predict the health of plant and also in this SVM(SUPPORT VECTOR MACHINE) classification algorithm is used for disease prediction. In [4] Rajesh et.al. provides information on classification technique which is used to detect disease in plant leaves as well as algorithm for image partition technique which is used for detection and classification of disease.In[6] Singh et.al. used various image processing technique to get different and useful features required for analyzing purpose.In[13]Barbedo et.al. used digital image based algorithm for automatic detection of plant disease. This algorithm deal with several disease . It is histogram based structure that makes it reasonably robust to the condition under which the images were captured.In[14]Muhammad et.al. has shown the conversion from RGB to CEILAB improves probability of salient region detection and extraction in addition to decreasing the time by reducing the number of iterations for convergence.In[17]Ramesh et.al used Random forest classifier model for training the dataset and HOG for feature extraction.

## III. METHODOLOGY

### 1) Dataset Included:

- Plant Village

The dataset used in our proposed study has been taken from PlantVillage-Dataset which was collected from kaggle[19].The data used for modeling is of Apple leaves.For training purpose the dataset comprises of 2 folder named diseased and healthy which contain the images of leaves with respective labels.The diseased folder contains unhealthy leaves affected by Apple Scab,Black Rot or Cedar Apple Rust .The healthy folder consist of healthy images.Fig.1 shows the images of diseased leaf consist of apple scab,black rot and cedar apple rust.

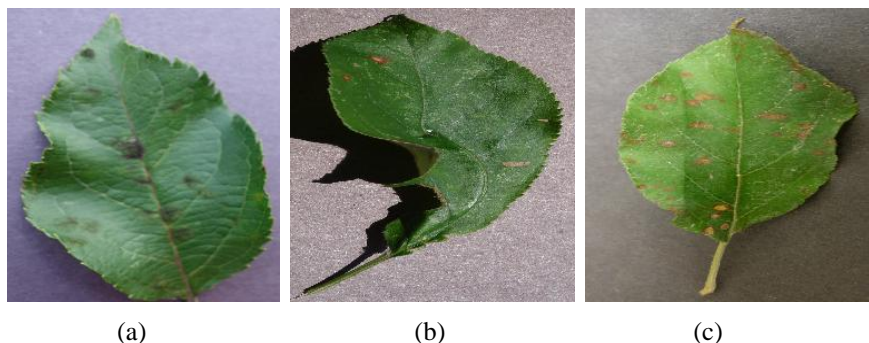


Fig.1.(a)Apple Scab( b)Black Rot (c)Cedar Apple Rust

## IV. PROPOSED MODEL

To identify whether the leaf is healthy or unhealthy following steps must be followed:1)Image preprocessing 2)Image Segmentation 3)Feature Extraction 4)Training of classifier and classification.

Preprocessing of image is done to improve the image quality and converting all the image size into a reduced uniform size for further processing.Image segmentation is done to separate the image of leaf from the background. Then the global feature descriptor is applied.Global features describes the image as whole to identify the entire object.Global feature are extracted from the preprocessed image using three descriptors i.e i)Hu moments ii)Haralick Texture iii)Color histogram.

i)Hu moments: Hu moments basically helps in describing the outline of a leaf. To calculate hu moment RGB is converted into Gray scale .

ii)Haralick Texture: Haralick Texture is used to distinguish between textures of healthy and unhealthy leaf as both of them usually has different texture.To calculate haralick texture image must be converted into grayscale.

iii)Color histogram: Color histogram represents the color of the image. In this RGB is first converted into HSV color and then histogram is calculated.Its important to convert RGB into HSV as HSV model relates closely with how human eye recognize the color in an image.

- Random Forest Classifier:  
Algorithm is implemented using random forest classifier.It is used for classification and regression technique. As compared to other machine learning algorithm like Support vector machine, Logistic regression, Navie bayes,K-nearest neighbour, Decision tree, Linear discriminant analysis,Random forest classifier is more flexible and give more accuracy with less number of dataset .Following is the architecture of proposed algorithm:

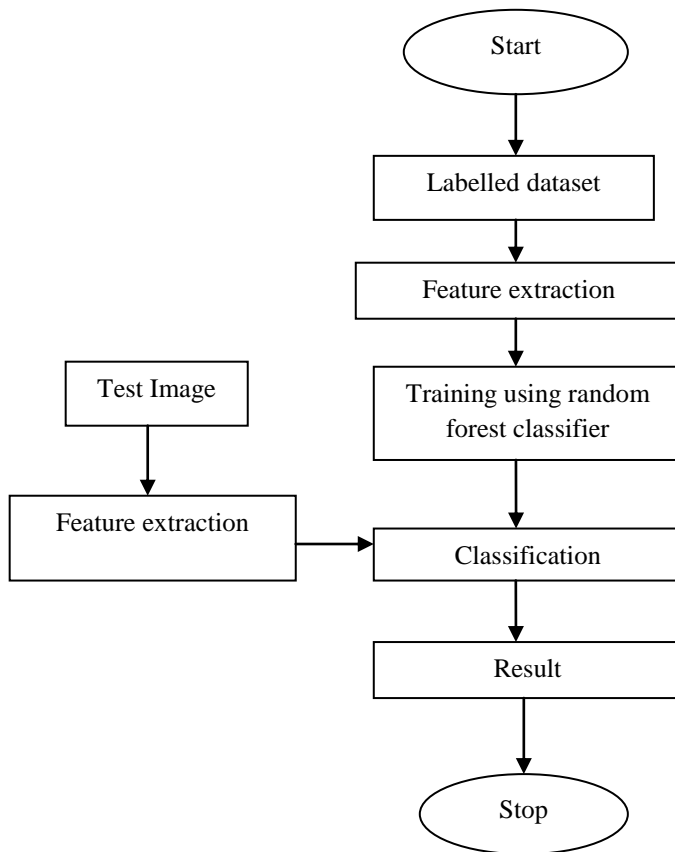


Fig.2. Flow chart of the proposed model

The dataset were divided into training and testing set with the ratio of 80/20 respectively.Feature vector is generated for training dataset by applying global feature extraction. Then generated feature vector is trained using random forest classifier.Then the feature vector for testing data generated using global feature extraction is given to trained classifier for predicting the result as referred in fig1.



V. EXPERIMENTAL RESULT

For calculating Hu moment and Haralick feature RGB image was converted into grayscale. Basically this is done so that we can calculate Hu moment descriptor and Haralick Texture over single channel. So its important to convert RGB into gray scale before calculating the HU moment and haralick tecture. For calculating histogram RGB image is converted into HSV image. As HSV separates the luma from the chroma. Fig.2 shows the conversion of RGB image into HSV.

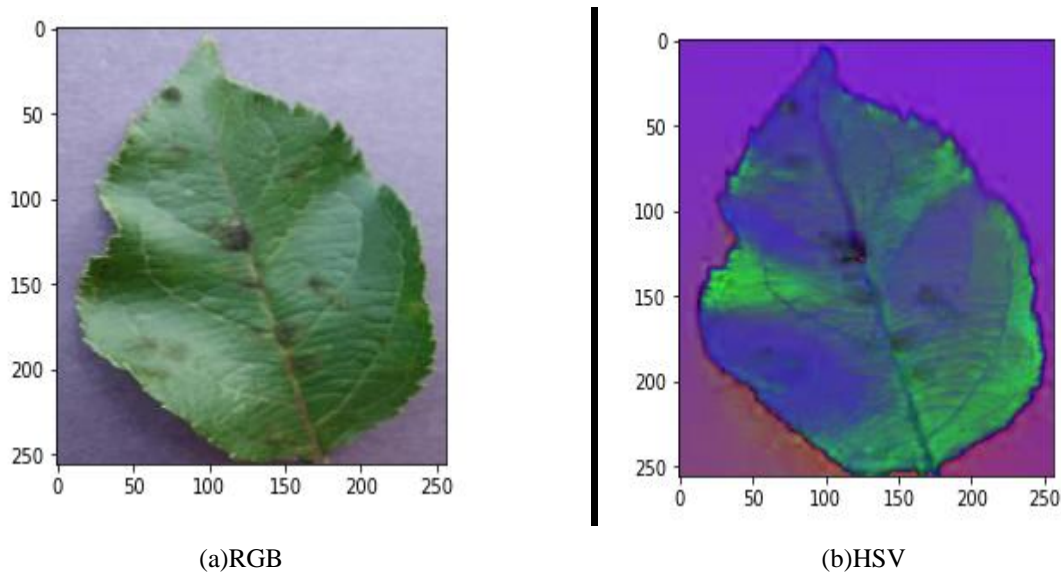


Fig.3. Conversion to HSV image format from RGB.

Machine Learning algorithm comparison

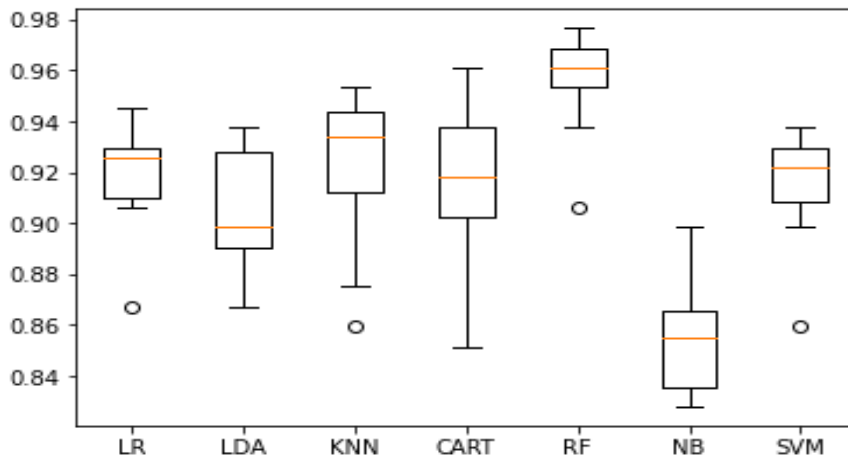


Fig.4. Comparison between different machine learning model.

As shown the in Fig.4 an accuracy of 98% is achieved using Random forest classifier which is more as compared to other machine learning models. Also the proposed model that we implemented gave an average precision-recall score of 0.98. Fig.4 shows the table based representation of the accuracy of different machine learning model. Fig.5 shows the confusion matrix of random forest classifier.



Table

Machine learning Model	Accuracy(Percent)
LR	91.95%
KNN	92.26%
Decision Tree	91.71%
Random Forest Classifier(Proposed Model)	98.12%
Naïve Bayes	85.54%
SVM	91.56%
LDA	90.46%

Fig.5.Showing the comparison between the different machine learning model

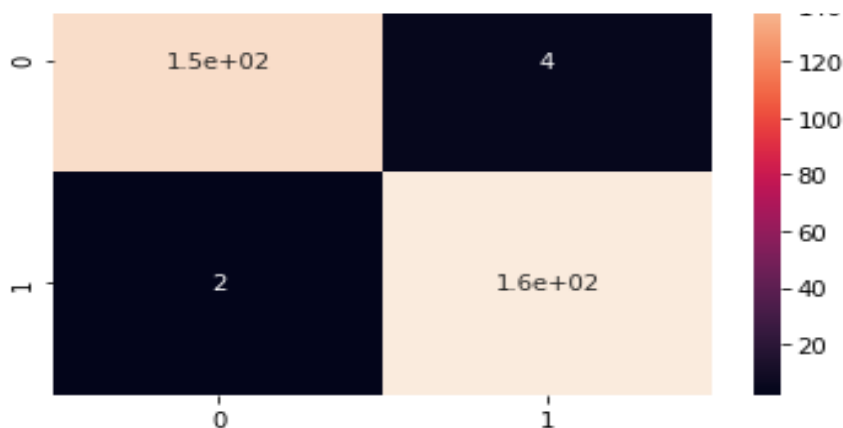


Fig.6. Confusion Matrix of Random forest classifier.

### VI. CONCLUSION

This proposed model is used to identify the abnormalities that occur on plant. This model was compared with other machine learning model for accuracy. Using Random forest classifier, labeled image based dataset is trained. The model gave the accuracy of 98%.This proposed model perform better then other machine learning model such as SVM,KNN, Logistic regression, NB, Decision Tree, LDA. So in future Random forest classifier will help for the rapid and effective identification of diseased leaves.

### VII. ACKNOWLEDGEMENT

I would like to thank Prof. Swapna Augustine Nikale, Department of Information Technology, B.K Birla College Kalyan for guiding throughout the research work.

### VIII. GLOSSARY

**LR**-Logistic Regression

**KNN**-K-Nearest Neighbour

**SVM**-Support Vector Machine

**LDA**-Linear Discriminant Analysis

**HSV**-Hue,Saturation,Value

## REFERENCES

- [1] Devaraj, A., Rathan, K., Jaahnavi, S., & Indira, K. (2019). Identification of Plant Disease using Image Processing Technique. 2019 International Conference on Communication and Signal Processing (ICCSP), 1–5. <https://doi.org/10.1109/iccsp.2019.8698056>
- [2] Mwebaze, E., & Owomugisha, G. (2016). Machine Learning for Plant Disease Incidence and Severity Measurements from Leaf Images. 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA), 1–6. <https://doi.org/10.1109/icmla.2016.0034>
- [3] Jadhav, P. A., Sardeshmukh, P., Ghanvat, H., Kadale, R., & Suryawanshi, R. (2019). Plant Health Prediction Using Machine Learning. *JOURNAL OF ARCHITECTURE & TECHNOLOGY*, XI(V), 1–6. <http://xajzkjdx.cn/gallery/4-may2019.pdf>
- [4] Rajesh, B., Sai Vardhan, M. V., & Sujihelen, L. (2020). Leaf Disease Detection and Classification by Decision Tree. 2020th International Conference on Trends in Electronics and Informatics (ICOEI)(48184), 1–4. <https://doi.org/10.1109/icoei48184.2020.9142988>
- [5] Singh, V., & Misra, A. K. (2017). Detection of plant leaf diseases using image segmentation and soft computing techniques. *Information Processing in Agriculture*, 4(1), 41–49. <https://doi.org/10.1016/j.inpa.2016.10.005>
- [6] Singh, V., Varsha, & Misra, A. K. (2015). Detection of unhealthy region of plant leaves using image processing and genetic algorithm. 2015 International Conference on Advances in Computer Engineering and Applications, 1–5. <https://doi.org/10.1109/icacea.2015.7164858>
- [7] Elangovan, K., & Nalini, S. (2017). Plant Disease Classification Using Image Segmentation and SVM Techniques. *International Journal of Computational Intelligence Research*, 13, 1–8. [https://www.ripublication.com/ijcir17/ijcirv13n7\\_23.pdf](https://www.ripublication.com/ijcir17/ijcirv13n7_23.pdf)
- [8] Moghadam, P., Ward, D., Goan, E., Jayawardena, S., Sikka, P., & Hernandez, E. (2017). Plant Disease Detection Using Hyperspectral Imaging. 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 1–8. <https://doi.org/10.1109/dicta.2017.8227476>
- [9] Mahlein, A.-K. (2016). Plant Disease Detection by Imaging Sensors – Parallels and Specific Demands for Precision Agriculture and Plant Phenotyping. *Plant Disease*, 100(2), 241–251. <https://doi.org/10.1094/pdis-03-15-0340-fe>
- [10] Archana, K. S., & Sahayadhas, A. (2018). Automatic Rice Leaf Disease Segmentation Using Image Processing Techniques. *International Journal of Engineering & Technology*, 1–4. <https://dl.wqtxts1xzle7.cloudfront.net/60753757/segmentationpaper120190930-127907-roknd4.pdf>
- [11] Kajale, R. (2015). DETECTION & RECOGNIZATION OF PLANT LEAF DISEASES USING IMAGE PROCESSING AND ANDROID O.S. *International Journal of Engineering Research and General Science.*, 3(2), 1–4. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.735&rep=rep1&type=pdf>
- [12] Rath, A. K., & Meher, J. K. (2019). CC. *Archives of Phytopathology and Plant Protection*, 52(19–20), 1348–1358. <https://doi.org/10.1080/03235408.2019.1708546>
- [13] Barbedo, J. G. A., Koenigkan, L. V., & Santos, T. T. (2016). Identifying multiple plant diseases using digital image processing. *Biosystems Engineering*, 147, 104–116. <https://doi.org/10.1016/j.biosystemseng.2016.03.012>
- [14] Khan, Z. U., Akram, T., Naqvi, S. R., Haider, S. A., Kamran, M., & Muhammad, N. (2018). Automatic detection of plant diseases; utilizing an unsupervised cascaded design. 2018 15th International Bhurban Conference on Applied Sciences and Technology (IBCAST), 1–8. <https://doi.org/10.1109/ibcast.2018.8312246>
- [15] Kumar, S. S., & Raghavendra, B. K. (2019). Diseases Detection of Various Plant Leaf Using Image Processing Techniques: A Review. 2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS), 1–4. <https://doi.org/10.1109/icaccs.2019.8728325>
- [16] Patil, R., Udgave, S., More, S., Nemishte, D., & Kasture, M. (2016). Grape Leaf Disease Detection Using K-means Clustering Algorithm. *International Research Journal of Engineering and Technology (IRJET)*, 1–4. <https://dl.wqtxts1xzle7.cloudfront.net/54619653/IRJET-V3I4461.pdf>
- [17] Ramesh, S., Hebbar, R., M., N., R., P., N., P. B., N., S., & P.V., V. (2018). Plant Disease Detection Using Machine Learning. 2018 International Conference on Design Innovations for 3Cs Compute Communicate Control (ICDI3C), 1–5. <https://doi.org/10.1109/icdi3c.2018.00017>
- [18] Samajpati, B. J., & Degadwala, S. D. (2016). Hybrid approach for apple fruit diseases detection and classification using random forest classifier. 2016 International Conference on Communication and Signal Processing (ICCSP), 1–5. <https://doi.org/10.1109/iccsp.2016.7754302>
- [19] <https://www.kaggle.com/abdallahalidev/plantvillage-dataset/version/1>



**INNO**  **SPACE**  
SJIF Scientific Journal Impact Factor

**Impact Factor:  
7.512**

**ISSN** INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
**INDIA**



# INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 **9940 572 462**  **6381 907 438**  **ijirset@gmail.com**



[www.ijirset.com](http://www.ijirset.com)

Scan to save the contact details