**Plant Disease Detection Using Image Processing and**

**Machine Learning**

**Abstract.** One of the important and tedious task in agricultural practices is detection of disease on crops. It requires huge time as well as skilled labor. This paper proposes a smart and efficient technique for detection of crop disease which uses computer vision and machine learning techniques. The proposed system is able to detect 20 different diseases of 5 common plants with 93% accuracy.

**Keywords:** Digital image processing, Foreground detection, Machine learning, Plant disease detection.

* **Introduction**

In India about 70% of the populace relies on agriculture. Identification of the plant diseases is important in order to prevent the losses within the yield. It's terribly troublesome to observe the plant diseases manually. It needs tremendous quantity of labor, expertize within the plant diseases, and conjointly need the excessive time interval. Hence, image processing and machine learning models can be employed for the detection of plant diseases. In this project, we have described the technique for the detection of plant diseases with the help of their leaves pictures. Image processing is a branch of signal processing which can extract the image properties or useful information from the image. Machine learning is a sub part of artificial intelligence which works automatically or give instructions to do a particular task. The main aim of machine learning is to understand the training data and fit that training data into models that should be useful to the people. So it can assist in good decisions making and predicting the correct output using the large amount of training data. The color of leaves, amount of damage to leaves, area of the leaf, texture parameters are used for classification. In this project we have analyzed different image parameters or features to identifying different plant leaves diseases to achieve the best accuracy. Previously plant disease detection is done by visual inspection of the leaves or some chemical processes by experts. For doing so, a large team of experts as well as continuous observation of plant is needed, which costs high when we do with large farms. In such conditions, the recommended system proves to be helpful in monitoring large fields of crops. Automatic detection of the diseases by simply seeing the symptoms on the plant leaves makes it easier as well as cheaper. The proposed solution for plant disease detection is computationally less expensive and requires less time for prediction than other deep learning based approaches since it uses statistical machine learning and image processing algorithm.

* **Methodology and Working**
* **Dataset**

For this project we have used public dataset for plant leaf disease detection called PlantVillage curated by Sharada P. Mohanty et Al. [6]. The dataset consists of 87000 RGB images of healthy and unhealthy plant leaves having 38 classes out of which We have selected only 25 classes for experimentation of our algorithm These classes are shown in Table 1.

**Table 1.** Dataset Specifications.

|  |  |  |
| --- | --- | --- |
| Plant | Disease Name | No. of Images |
| Apple | Healthy Diseased Scab  Diseased: Black rot Diseased: Cedar apple rust | 2008  2016  1987  1760 |
| Corn | Healthy  Diseased: Cercospora leaf spot  Diseased: Common rust Diseased: Northern Leaf Blight | 1859  1642  1907  1908 |
| Grapes | Healthy  Diseased: Black rot  Diseased: Esca (Black Measles) Diseased: Leaf blight (Isariopsis) | 1692  1888  1920  1722 |
| Potato | Healthy  Diseased: Early blight Diseased: Late blight | 1824  1939  1939 |
| Tomato | Healthy  Diseased: Bacterial spot Diseased: Early blight Diseased: Late blight Diseased: Leaf Mold Diseased: Septoria leaf spot  Diseased: Two-spotted spider mite Diseased: Target Spot  Diseased: Yellow Leaf Curl Virus Diseased: Tomato mosaic virus | 1926  1702  1920  1851  1882  1745  1741  1827  1961  1790 |

Some samples from the datset are shown in Fig. 1.



**Fig. 1.** Sample images in the dataset [6].

* **Data preprocessing and feature extraction**

Data preprocessing is important task in any computer vision based system. Fig. 2 illustrates the preprocessing steps for each image. To get precise results, some background noise should be removed before extraction of features. So first the RGB image is converted to greyscale and then Gaussian filter is used for smoothening of the image. Then to binaries the image, Otsu’s thresholding algorithm is implemented. Then morphological transform is applied on binarised image to close the small holes in the foreground part. Now after foreground detection, the bitwise AND operation on binarised image and original color image is performed to get RGB image of segmented leaf. Now after image segmentation shape, texture and color features are extracted from the image. By using contours, area of the leaf and perimeter of the leaf is calculated. Contours are the line that joins all the points along the edges of objects having same color or intensity. Mean and standard deviation of each channel in RGB image is also estimated. To obtain amount of green color in the image, image is first converted to HSV color space and we have calculated the ratio of number of pixels having pixel intensity of hue (H) channel in between 30 and 70 and total number of pixels in one channel. Non green part of image is calculated by subtracting green color part from 1.

After extracting color features from the image, we have extracted texture features from grey level co-occurrence matrix (GLCM) of the image [7].

**Fig. 2.** Steps for data preprocessing and feature extraction.

GLCM is the spacial relationship of pixels in the image. Extracting texture features from GCLM is one of the tradition method in computer vision. We have extracted following features from GCLM:

* Contrast
* Dissimilarity
* Homogeneity
* Energy
* Correlation

After extracting all the features from all the images in the dataset, feature selection task is performed.

* **Feature selection**

Feature selection is an important step in all machine learning problems. In this project we are selecting the features on the basis of correlation of variables with target variable. Fig. 3 shows the correlation of each variable with each other for apple dataset. The correlation of feature green part of leaf (F1) and green part of leaf (F2) is very high (1) which means both variables are dependent on each other. So we have dropped one of them (F2). Now for apple disease prediction, less correlated features such as green channel mean, red channel standard deviation, blue channel standard deviation, dissimilarity (f5) and correlation (f8) will not contribute too much in model development. So we have dropped these variables also. After feature selection, the data is now parsed to machine learning classifiers to find the patterns in the data.

**Fig. 3.** Correlation plot for Apple dataset.

* **Classification Algorithm**

Random forest classifier has been used for classification or detection task. It is the part of ensemble learning, where the output is predicted from multiple base estimators [8]. Generally, to achieve higher accuracies, decision trees are used. But they are prone to overfitting problems. So to overcome this issue, random forest classifier is used which is a combination of multiple decision trees. Each tree is trained by using different subsets of the whole dataset, this can reduce the overfitting and improves the accuracy of the classifier. We have splitted the dataset into train set (80%) for fitting the model and test set (20%) for validation. K-fold cross validation technique is implemented to find the accuracy score. This method can find the accuracy on whole dataset without any bias. After fitting the data, f1 score, precision, recall, accuracy has been calculated

from test data to analyze the performance of the model. ROC curve and confusion matrix was plotted to analyze false positives and false negatives.

* **Results and discussion**

Table 2 shows the performance matrices for each model developed for each of the plant. We can observe that the accuracy scores are nearly equal to f1 scores. This is because of balanced number of false negative and false positive predictions. This is considered as best case for any machine learning algorithm. The average accuracy was 93%.

**Table 2.** Performance matric for all models.

|  |  |  |
| --- | --- | --- |
| Plant | Accuracy | F1 Score |
| Apple Corn Grapes Potato | 0.91  0.94  0.95  0.98 | 0.91  0.94  0.95  0.98 |
| Tomato | 0.87 | 0.87 |



**Fig. 5.** Images and outputs generated by system.

* **Conclusion**

We have successfully developed a computer vision based system for plant disease detection with average 93% accuracy and 0.93 F1 score. Also the proposed system is computationally efficient because of the use of statistical image processing and machine learning model. Table 3 illustrates the overall benefits of our system over the other approaches.

**Table 3.** Comparison of proposed system with other existing systems.



We can observe that our technique is accurate and efficient compared with other systems. Also it won't require a specialized hardware, makes it cost effective solution.

**References**

* S. D. Khirade and A. B. Patil, "Plant Disease Detection Using Image Processing," 2015 International Conference on Computing Communication Control and Automation, 2015, pp. 768-771, doi: 10.1109/ICCUBEA.2015.153.
* S. C. Madiwalar and M. V. Wyawahare, "Plant disease identification: A comparative study," 2017 International Conference on Data Management, Analytics and Innovation (ICDMAI), 2017, pp. 13-18, doi: 10.1109/ICDMAI.2017.8073478.
* P. Moghadam, D. Ward, E. Goan, S. Jayawardena, P. Sikka and E. Hernandez, "Plant Disease Detection Using Hyperspectral Imaging," 2017 International Conference on Digital Image Computing: Techniques and Applications (DICTA), 2017, pp. 1-8, doi: 10.1109/DICTA.2017.8227476.
* S. D.M., Akhilesh, S. A. Kumar, R. M.G. and P. C., "Image based Plant Disease Detection "

This Project Carried out by

S. Vasanth (TL)

V. Vijay Avinash

C. Manimaran

T. Balaji