JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Experimentation with Plant Disease Detection Using ML

Mr.Atharva Santosh Bhagwat,Mr.Tanmay Pramod Garde,Mr.Soham Uday Khedekar,

Mr.Rutik Ravindra Vaskar,Prof. Mandar S Joshi.

BE Information Technology,BE Information Technology,BE Information Technology
BE Information Technology,Professor
Information Technology,
Finolex Academy Of Management and Technology
, Ratnagiri,India

Abstract : Agriculture is an important industry in India. Most of the population depends on this industry, so the life of the crop produced is very important. This makes plant diseases the major enemy to this industry.
But with today's modern world it is easy to overcome these diseases with the help of Image processing. With image processing it becomes easier to accurately detect the exact disease that is affecting the plant. In this project we aim to create an application for the disease detection for plants using Image processing & machine learning. This paper goes in detail regarding the same. This includes various steps like image acquisition, image pre-processing, feature extraction.

IndexTerms **- Agriculture, plant diseases, image processing, image acquisition, image pre-processing, feature extraction.**

I. INTRODUCTION

India is an agricultural country & most of its economy is built due its agricultural industry. Most of the population depends on Agriculture. The goods produced are also exported out to the whole world which helps in a good economy and relationships with other countries. Hence the goods produced must be of top quality & produced in high quantity. But plant diseases can bring a halt to this massive industry leading to a lot of problems.

Disease generally mean illness of people, animals, plants, etc., caused by infection or a failure of health [1]. In plants, diseases are caused by various pathogens, bacteria, fungi, etc. This disease can affect various vital functions in plants that result in their inability to produce the desired results. So it is necessary to identify these diseases as soon as possible to prevent any major loss. Lot of times, even experienced farmers find it difficult to identify the disease as the symptoms of many diseases can look very similar. Getting expert help would be expensive & a long process. This is where image processing comes in handy.

With image processing & Machine learning, it will be easy, quick & efficient to identify & distinguish between these diseases before they can cause some major damage. As this will be machine work, the accuracy of the results will be high. The focus of this paper is to concentrate on Plant leaves to detect the texture of the leaf & detect the disease.

This paper is organized into following sections. Section 1 is the introductory part on plant diseases, importance of disease detection. Section 2 is a literature review on the different research papers that were studied before working on the projects. Section 3 of this paper tackles methodology i.e a detailed report on how the said system works & Finally Section 4 is where the paper is concluded with the observed results of the system.

II. PURPOSE

1. To detect plant disease from images of leaves of a plant.
2. To build a model for disease detection using machine learning.

III. LITERATURE SURVEY

After reading and understanding the following research papers we devised the inference methodology to solve this problem which is stated in following table:

Research Papers	Inference
An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques Ms. Kiran R. Gavhale1, Prof. Ujwalla Gawande2 [1]	Methodology 1) RGB image acquisition. 2) convert the input image into color space. 3) Segment the components. 4) Obtain useful segments. 5) Computing the texture features. 6) Configure the neural networks for recognition. Model trained on Real-world Data perform better than lab created data for application in real-world Algorithms: CNN Random Forest SVM
L. Li et al.: Plant Disease Detection and Classification by Deep Learning—A Review [2]	
Using Deep Learning for Image-Based Plant Disease Detection [3]	
Plant Disease Detection Using Image Processing and Machine Learning [4]	
PlantDoc: A Dataset for Visual Plant Disease Detection [5]	

IV. PLANT DISEASE CONSIDERATE

To identify the diseases in different plants and learn about features we are considering following plants for the study:

Plant	Plant Disease
Tomato Plant	Tomato Bacterial spot Tomato Early blight Tomato Late blight
Potato Plant	Potato Early blight Late blight
Pepper bell Plant	Pepper bell bacterial spot

Tomato Healthy:


Tomato Bacterial spot:


Tomato Early blight :


Tomato Late blight :


Potato Healthy:


Potato Early blight :


Potato Late blight :


Pepper bell Healthy:


Pepper bell bacterial spot :

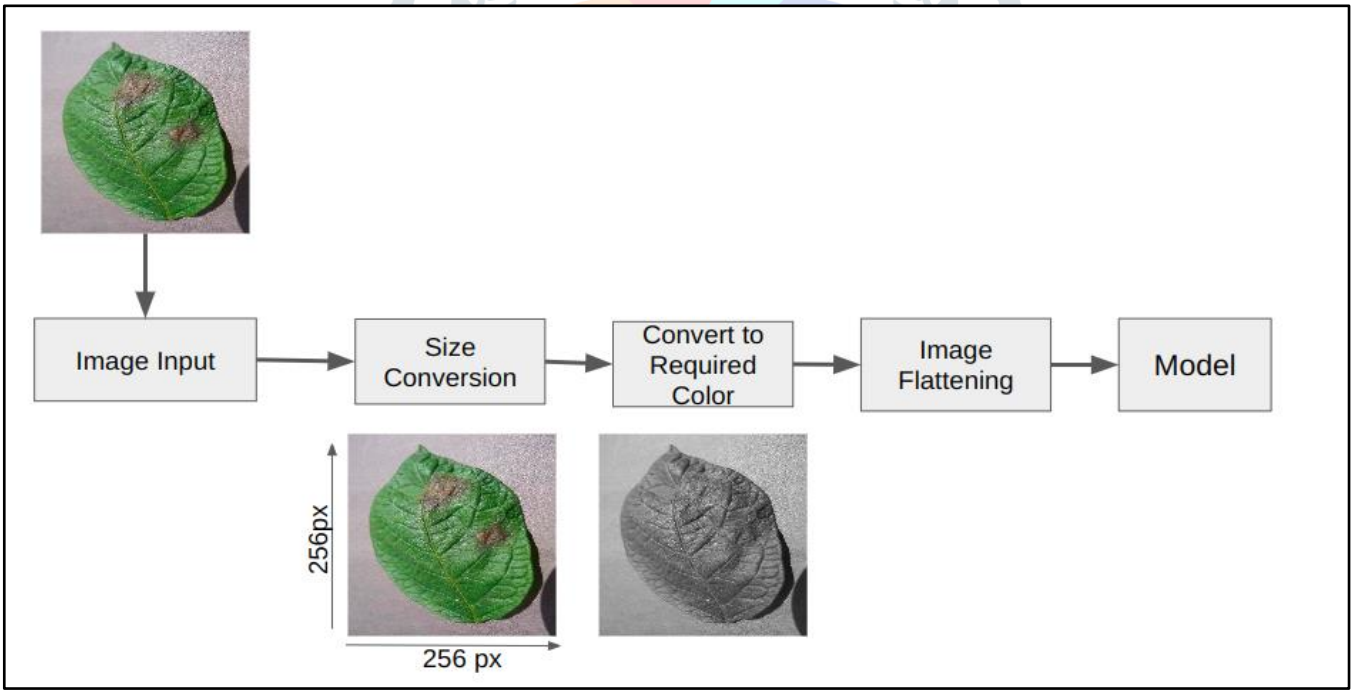

V. RESEARCH METHODOLOGY

A model is trained to detect the target classes (things to identify in images) using labeled example images. Image classification is a supervised learning problem. The input for early computer vision models was simply raw pixel data.
In this project we are going to classify images of leaves of plant to its corresponding disease
For this we are using Random Forest Machine learning algorithm.Random Forest is a well-known machine learning algorithm from the supervised learning technique. It can be applied to ML problems involving both classification and regression. It is built on the idea of ensemble learning, which is a method of integrating various classifiers to address difficult issues and enhance model performance.
Random Forest, as the name implies, is a classifier that uses a number of decision trees on different subsets of the provided dataset and averages them to increase the dataset's predictive accuracy. Instead than depending on a single decision tree, the random forest uses forecasts from each tree and predicts the result based on the votes of the majority of predictions.
The greater number of trees in the forest leads to higher accuracy and prevents the problem of overfitting.
For solving this we are building a model based on max voting ensemble method which aggregates results from all models and gives results accordingly.
The models are trained on images of plant leaves with different diseases which are converted to 256x256 size and them transformed to R,G,B and Gray scale to feed these images to respective models which are build for each transformed type of image input data, the aggregator aggregate these results and gives more accurate results.
The input images can be converted to a higher size than 256x256,512x512 etc. which may result in better performance but it is not implemented due hardware limitations.
Image Preprocessing

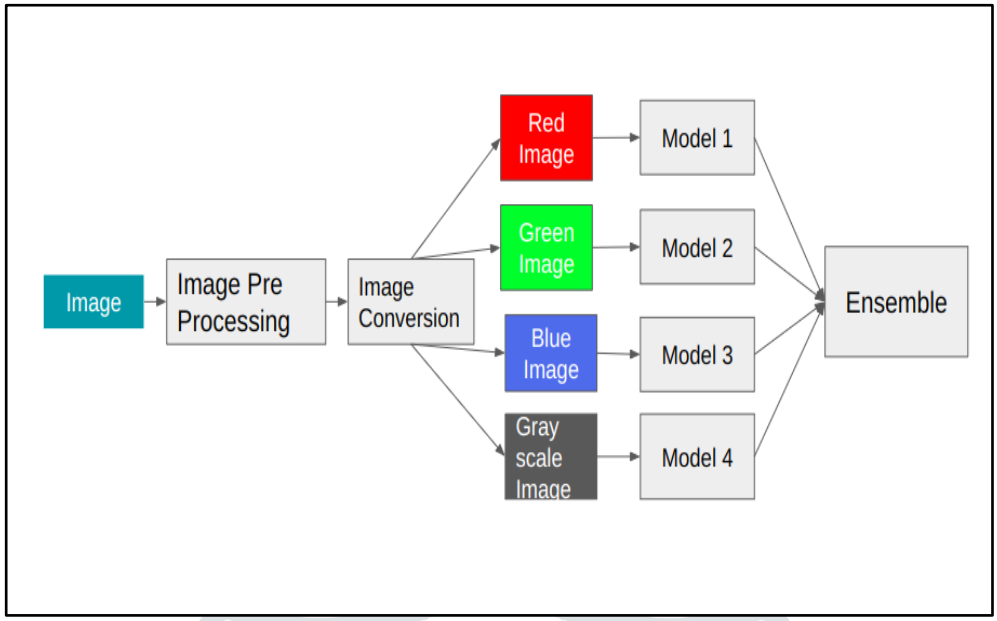
After image acquisition in image preprocessing the size conversion of the input image from original size to 256pxx256px.Then the image is converted to red,blue ,green and grayscale image.This image is flattened in image flattening to feed it to Machine Learning Model.The flattening helps in training of model by reducing dimensionality thus less load while training.

VI. Image Preprocessing

After image acquisition in image preprocessing the size conversion of the input image from original size to 256pxx256px.Then the image is converted to red,blue ,green and grayscale image.This image is flattened in image flattening to feed it to Machine Learning Model.The flattening helps in training of model by reducing dimensionality thus less load while training.



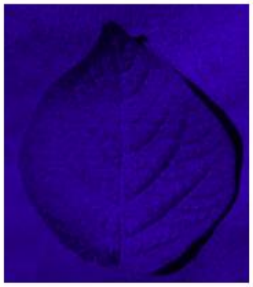
VII. PROPOSED SYSTEM



The proposed system takes image as input and in preprocessing transform them into 4 forms of image namely red,green,blue and gray and gives it models for training which fit images to respective model the ensembler/aggregator generates result on basis of maximum voting.

The below table shows the preprocessed images of potato plant in following order:

1)Potato Healthy
2)Potato Early Blight
3)Potato Late Blight

Gray Scale	Red	Blue	Green
			
			
			

VIII. RESULTS

The Images of Tomato Plant were trained on Random Forest Classifiers with different combinations of Estimators , Training Data Sample Size and The Color of the pixel used for Training and Testing . The following are the Results of this testing -

Type of Image - Grayscale

Sr No	Training Data Size	Estimators (in Random forest)	Testing Data Size	Accuracy
1)	200	4	80	70 %
2)	200	7	100	80%
3)	800	3	100	79%
4)	800	7	100	82%
5)	800	7	200	86.5%
6)	800	11	200	89.5%

Type of Image - Red

Sr No	Training Data Size	Estimators (in Random forest)	Testing Data Size	Accuracy
1)	200	4	200	71.5%
2)	200	7	100	71%
3)	800	3	100	81%
4)	800	7	100	84%
5)	800	7	200	85.5%
6)	800	11	200	88.5%

Type of Image - Green

Sr No	Training Data Size	Estimators (in Random forest)	Testing Data Size	Accuracy
1)	200	4	100	70%
2)	200	7	100	80%
3)	800	3	100	77%
4)	800	7	100	81%
5)	800	7	200	85%
6)	800	11	200	86%

Type of Image - Blue

Sr No	Training Data Size	Estimators (in Random forest)	Testing Data Size	Accuracy
1)	200	4	100	81%
2)	200	7	100	90%
3)	800	3	100	89%
4)	800	7	100	92%
5)	800	7	200	91%
6)	800	11	200	91.5%

Type of Image - Ensemble

(Combined results of Red,Green,Blue and Grayscale)

Sr No	Training Data Size	Estimators (in Random forest)	Testing Data Size	Accuracy
1)	1200	7	200	92%
2)	1200	11	200	94.5%

***Accuray may have variations of upto 2% due to random sampling used in Random forest classifier**

IX. INFERENCE

1) The Performance of Image Classification is directly proportional to the size of training data and the number of estimators used in random forest classifiers.
2) In case of Grayscale smaller training data with more estimators gives accuray similar to larger training data and less number of estimators .
3) In case of Grayscale , Red , Green Images increasing estimators from 7 to 11 Increases the accuracy but the Accuracy does not significantly increase in case of Blue Images
4) Comparing the results of Red,Blue,Green and Grayscale images used for training , the performance of each Image type is in descending order is -
   Blue , Grayscale , Red ,Green
5) The Ensemble model is created by Aggregating the results of Red,Blue,Green and Grayscale Images on respective models and then Selecting the Output based on Maximum Voting .
6) The Ensemble model provides the highest Accuracy in classifying the images .

REFERENCES

[1] An Overview of the Research on Plant Leaves Disease detection using Image Processing Techniques Ms. Kiran R. Gavhale1, Prof. Ujwalla Gawande2
[2] L. Li et al.: Plant Disease Detection and Classification by Deep Learning—A Review
[3] Using Deep Learning for Image-Based Plant Disease Detection
[4] Plant Disease Detection Using Image Processing and Machine Learning
[5] PlantDoc: A Dataset for Visual Plant Disease Detection Davinder Singh*, Naman Jain*, Pranjali Jain*, Pratik Kayal* ,Sudhakar Kumawat, Nipun Batra