IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Plant Disease Detection Using Image processing and Machine learning Techniques

Jamil Ahemad Abdul Hafiz Shaikh
M.E Student, Department of Computer Engineering, MCOER, Nashik.

***Abstract*—The prevention and control in plant disease plays vital role in architectural filed. Accurate and rapid diagnostic of disease helps to control the disease at early stage. The detection of plant disease using automatic technique is beneficial and reduces large work of monitoring each individual plant in the farm. A combined usage of image processing techniques and machine learning techniques helps to recognize the disease. In the proposed system, plant leaf image features are extracted using Gabor filter and watershed segmentation algorithm which includes the color, texture and intensity property of image. Based on the extracted features the image is compared with existing disease training data. The test image is labeled with the related disease using classification technique. The proposed study focuses on the comparative analysis using various classification techniques analysis in disease detection.**

Keywords*— *Image Processing, Machine learning, Plant Disease, Neive Bays, SVM, Random Forest

I. INTRODUCTION

In India the architectural productivity plays vital role in economy. Variety of plants and crops present in India. The plants have variety of characteristics and behavior. The Plant disease affects the productivity. The prevention and control in plant disease plays vital role in architectural filed. Accurate and rapid diagnostic of plant disease helps to control the disease at early stage.

The disease detection in plants plays important role in architectural field. The automatic detection of plant disease using automatic technique is beneficial and reduces large work of monitoring each individual plant in the farm.

The plant disease are mainly categorized in two sections: Biotic and Abiotic. Biotic diseases affect the plant stem and leaf. The Abiotic disease occurs due to unhealthy soil, wind and it affects root. For Biotic plant disease classification the plant leaf images are analyzed. The image features are extracted using feature selection algorithm using image processing technique. These features are matched with existing training dataset using machine learning classification technique. This technique assigns the class label for image based on the training information provided.

Analysis of Different Classification Techniques using machine learning is a challenging task in agricultural research area. It is very tough to find out the best classification algorithms for comparing in different algorithms in various datasets. The proposed work concerns with the algorithm and its capability to diagnose plant disease data accurately as well as quickly.

The proposed system works on plant disease detection using image processing and machine learning techniques. The plant leaf image is taken for processing and identification of disease. The segmentation is treated as preprocessing step. Then the system extracts image features in terms of color, texture and intensity. Further the data is provided to machine learning classifiers to categorize and correctly label the plant disease. For classification Random Forest, Neive bays and SVM algorithms are used.

II. RELATED WORK

S. Avinash .et.al.[2], proposes an improved image processing analysis for the detection of lung cancer. To overcome the drawback of analysis of only color and texture filters the watershed segmentation algorithm is combined with color and texture filters to improve the accuracy of detection. This paper proposes a lung cancer detection using image processing and techniques. It uses Gabor filter followed by watershed algorithm and then masking technique to correctly map the area. This technique does not use any classification algorithm for detection.

Vijai Singh, et. al.[3] proposes a technique for Detection of plant leaf diseases using image segmentation and soft computing techniques. This paper also focuses on plant disease detection using monitoring plant leaf. For labeling the correct disease it uses image processing and machine learning technique. For feature extraction it uses Segmentation algorithm whereas for classification it uses SVM classifier. In the future work of this paper state that, to improve classification accuracy hybrid approach of multiple classification algorithms like neive bays, fuzzy logic can be used.

Hybrid Approach for Apple Fruit Diseases Detection is proposed by Bhavini J.[4]. This paper focuses on fruit disease detection technique. For detection it initially uses color and texture filters for extracted feature of fruit. For color and texture feature extraction it uses variety of combination of algorithms such as: GCH+ Gabor, GCH+LTP, Gabor+LTP, Gabor+CLBP+LTP, etc. After extraction of features random forest classification algorithm is used to correctly label the disease.

Comparative Analysis of classification Algorithms using WEKA tool is elaborated in the paper presented by Shivangi Gupta[5].There are various classification algorithms are proposed in machine learning. The weka tool gives the implementation of

some classification algorithms in java. This paper proposes a comparative analysis of various classification algorithms. The test is conducted on various datasets downloaded from UCI repository and accuracy of classification is calculated for each classifier.

Mr. N.S. Bharti [6] proposes a study on Detection and Classification of Plant Diseases. The system detects the plant diseases using in phases:1 :feature Extraction 2: Masking and 3: Classification. In feature extraction it uses segmentation. And Find the green part of image. In masking phase it uses Ostu's method to mask green pixel. After masking the green pixel, the pixels having zero RGB value and pixels on leaf boundaries are removed. Then the affected area is correctly cropped from the leaf. The extracted features of damaged area are then provided to the ANN classification algorithm and disease label is extracted.

R. Ravikumar[7] present a system that work on identification of the sugarcane plant leaf disease. using Feed Forward Artificial Neural Network technique and Kmeans algorithm. This technique uses Feed Forward Artificial Neural Network using Multilayer Perceptron (MLP) and Simple K means algorithm.

Milos Ilic[8] , et.al., proposes a technique for early fruit disease detection. This is disease prediction system. It uses previous plant disease data with various other parameters. For disease prediction it uses mathematical regression methods. This helps to suggest important chemical protection for plants.

R.Ramya , et. al. [1] proposes a cumulative study of various classifiers present in weka machine leaning tool for plant disease detection. The plant disease is identified using plant leaf image analysis. The features from images are extracted using color and texture based filters. For classification random forest, Zero R, Neive bays and SVM algorithms are used collectively to predict the result. Only color and texture filters do not appropriately find the disease in plant leaf using image processing.

III. ANALYSIS AND PROBLEM FORMULATION

Lot of machine learning techniques and image processing techniques are applied for disease detection and identification. These techniques are studied independently. To improve the system accuracy ensemble approach can be useful.

System should automatically detect the plant is suffering from disease or not. The disease should be correctly labeled by analyzing the plant leaf using image processing and machine learning techniques.

IV. PROPOSED METHODOLOGY

A. Architecture

Following figure shows the architecture of the system. The system takes training and testing images as an input. These are affected as well as normal plant leaf images. The system processes the image by applying feature extraction and then classification. Based on the classification result, system generates a label for image as affected or normal.

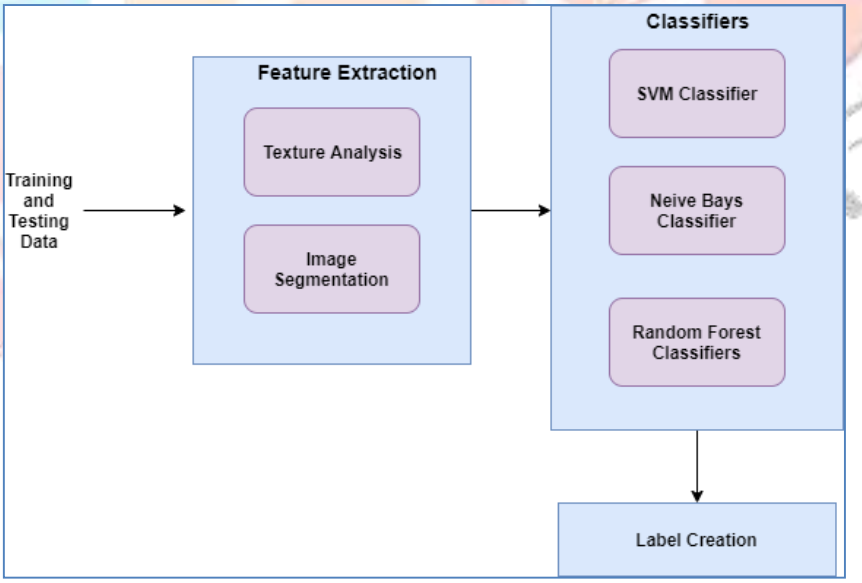

Fig. 1. System Architecture

B. System Working:

The proposed system contains following modules:

1. Feature Extraction:

Before applying the feature selection algorithms the image is resized and compressed to reduce the extra processing overhead.

In feature extraction color and texture based feature selection algorithms are applied on training dataset and on test image. The combination of color and texture filter algorithm is: gabor + LBP.

After color and texture feature extraction watershed segmentation algorithm is used. This separates the affected part of leaf. The feature extraction technique on training dataset is applied at ones and features of all images are extracted and saved with disease label. This saved data is used for classification training process.

1. Gabor Filter:

This is linear filter for texture analysis. It checks for specific frequency contents in the localized image region in specific direction. Mathematically it can be represented as:

$$g(x, y, \lambda, \theta, \Psi, \sigma, \Upsilon) = \exp\left(-\frac{x'^2 + \Upsilon^2 y'^2}{2\sigma^2}\right) \exp\left(i(2\pi\frac{x'}{\lambda} + \Psi)\right)$$

where,

λ = Wavelength of the sinusoidal component.

Θ =orientation of the normal to the parallel stripes of Gabor function.

Υ =spatial aspect ratio and specifies the ellipticity of the support of Gabor function.

Ψ = phase offset of the sinusoidal function.

σ =sigma/standard deviation of the Gaussian envelope

$x' = x.\cos\theta + y.\sin\theta$

$y' = x.\sin\theta + y.\sin\theta$

2. Watershed:

It is a segmentation method. It partitioned the image in multiple small pieces based on its properties. It works on grey scale image. It visualizes the image as topographic surface. High intensity pixels are treated as peaks whereas low intensity area is valleys. The watershed algorithm tries to fill valleys with water until all peaks are under water. To fill the valley section local minima of the image is used.

3. Classifier Training:

The SVM , Neive Bays and Random forest classifiers are used collectively to define a disease label. The feature extraction information of affected plant leaf is provided to these classifiers for training.

4. Classifier Testing:

The test image features are provided to these classifiers to extract correct label. A collective voting is conducted based on the result of each classifier and a summary label is generated based results of all classifiers.

5. Algorithms

Input: Tr: Training Image Dataset with label.
Tst: Test Image
Output: Lbl: disease label
Processing:
/* Training */

i. For each image I in Tr images
ii. extract feature using
   - Gabor filter
   - LBP
   - watershed algorithm
iii. generate training dataset with image features and disease lable
iv. Apply training for classifiers
   - Neive bays,
   - Random forest and
   - SVM

/* Testing */

v. For test image t extract feature using
   - Gabor filter
   - LBP
   - watershed algorithm
vi. Provide feature set as test instance to classifiers
   - Neive bays,
   - Random forest and
   - SVM

   and get label
vii. Generate voting average from the result of each classifier
viii. Return label

V. RESULT AND ANALYSIS

The system is implemented in java using jdk 1.8. For image processing opencv 3.0 framework is used. For machine learning algorithms Weka library is used. The system is implemented and tested on windows environment with core i3 processor and 4 gb ram.

A. Dataset:

Kaggle[10] plant leaf dataset is downloaded. It contains 87K rgb images of healthy and diseased crop leaves. The data is categorized in 38 different classes. The dataset contains leaf of Blueberry, apple, corn grape,etc.

B. Performance Measures:

The system performance is measured in terms of :

1. Time:

Time required for processing is captured. For various sized dataset the time required for training and testing is captured.

2. F-Measure:

F-measure is calculated using total test classification results: It is calculated using precision and recall. The formula is given as:

$$\text{F-measure} = 2 * \frac{\text{Precision} * \text{recall}}{\text{Pricision} + \text{recall}} \quad (1)$$

Where,

Precision is calculated using following formula:

$$\text{Precision} = \frac{TP}{TP+FP} \quad (2)$$

Recall is calculated using following formula:

$$\text{Recall} = \frac{TP}{TP+FN} \quad (3)$$

Where TP = true positive rate

FP = False positive rate

FN= false Negative Rate

3. Accuracy:

The accuracy of plant disease detection is calculated using following formula:

$$\text{Accuracy} = \frac{\text{number of correctly classified}}{\text{total number of images for testing}} \quad (4)$$

C. Results*:*

1. Feature Extraction

The dataset is loaded and its features are extracted using Gabor and Watershed algorithm. Following figure contains 3 images. The first image is input to the system the second image is output of Gabor filter and the third image is output of Watershed algorithm.

Fig. 2. Feature Extraction Results

The features of image are saved at the backend file system. The dataset contains filename, feature values and class label. The class label is healthy and unhealthy. The features of input image are matched with training data using classification technique and voting is performed for each classifier result. The collective result decides the class label of an image.

Time and Accuracy Results:

Following table contains the classification results for apple leaf image dataset. The table contains the classification time, and accuracy of the system. The time required for processing proposed system is higher than the existing system because proposed system extract features using gabor watershed algorithm. The proposed system has higher accuracy than existing system.

Table 1. Time and accuracy analysis

Images	Classification Accuracy- Existing System	Classification Accuracy -Proposed System	Classification Time-Existing System(in Sec)	Classification Time-Proposed System(in Sec)
200	73.02	73.43	10.53	11.32
400	73.32	74.2	14.47	13.41
600	73.08	74.05	20.32	19.85
800	74.32	74.53	26.23	26.59
1000	74.32	75.67	30.75	30.71

Following graph shows the time comparison between existing and proposed system. The proposed system requires higher time as compared to the existing system due to the added extra feature extraction technique. The time required for processing increases as we increase the dataset size.

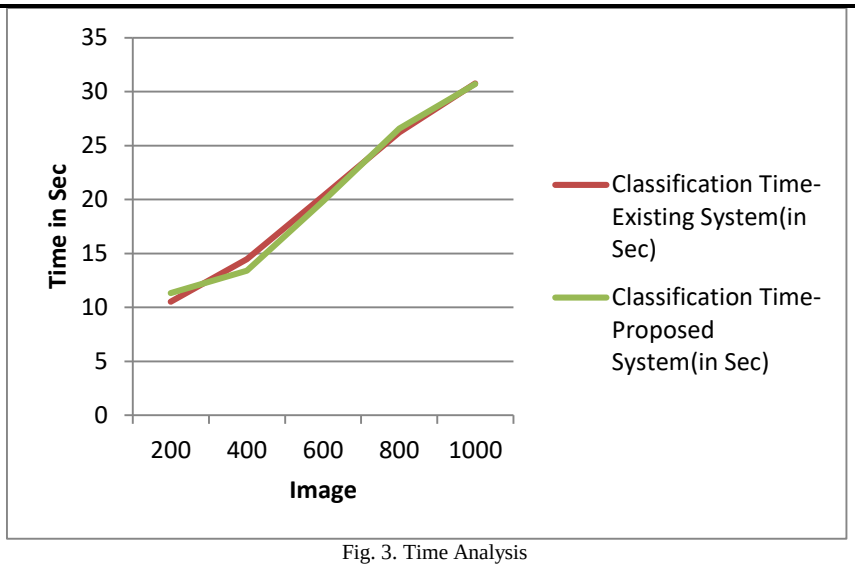

Fig. 3. Time Analysis

The following graph contains the accuracy comparison between existing and proposed system. The proposed system has higher accuracy as compared to the existing system. As we increase the training dataset size, accuracy also increases.

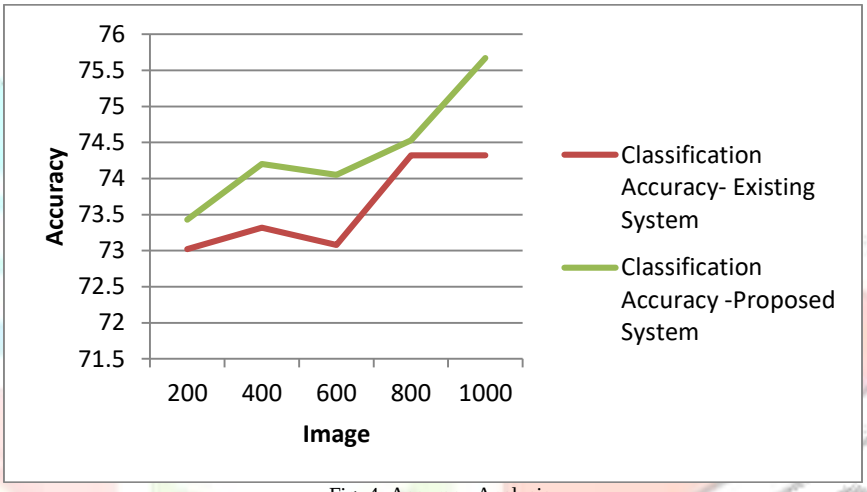

Fig. 4. Accuracy Analysis

VI. CONCLUSIONS

For automatic plant disease detection image processing and data mining techniques are used. Using image processing technique the features can be extracted of plant leaf or plant fruit. Based on the extracted features data mining technique conclude that plant is affected or not. The proposed system automatically detects the plant disease using image processing and machine learning algorithm. The plant leaf image features are extracted using image segmentation and texture analysis algorithm. The extracted features are compared with the training dataset with machine learning algorithms such as: SVM , Neive Bays and Random forest. The collective result of classification algorithm achieves the accuracy in detection. Based on the system execution following points are noticed:

- Better classification results are generated using watershed segmentation and gabor rather than only gabor filter
- The time required for processing is nearly equal for both feature extraction technique
- The collective result of classifier generates better results than individual classifier

In future system will work on automated plant disease detection using live hardware support.

REFERENCES

[1] R. Ramya, Dr.P.Kumar, D.Mugilan and M.Babykala, "A Review On Different classification Techniques In machine Learning using Weka for Plant disease Detection", in International Research Journal of Engineering and Technology, Vol:5,Issue :5, May 2018

[2] S. Avinash ; K. Manjunath ; S. Senthil Kumar, "An improved image processing analysis for the detection of lung cancer using Gabor filters and watershed segmentation technique", in International Conference on Inventive Computation Technologies (ICICT),INSPEC accn num: 16620663, Jan 2017

[3] Vijai Singh and A. K. Misra, "Detection of plant leaf diseases using image segmentation and soft computing techniques", in ScienceDirect Information Processing in Agriculture, Vol. 4, Issue 1, Pages 41-49, March 2017,

[4] Bhavini J. Samajpati and Sheshang D. Degadwala," Hybrid Approach for Apple Fruit Diseases Detection and Classification Using Random Forest Classifier" in International Conference on Communication and Signal Processing, April 6-8, 2016.

[5] Shivangi Gupta, Neeta Verma," Comparative Analysis of classification Algorithms using WEKA tool" International Journal of Scientific & Engineering Research, Volume 7, Issue 8, August-2016.

[6] Mr. N.S. Bharti, Prof. R.M. Mulajkar," Detection and Classification of Plant Diseases", in International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056 Volume: 02 Issue: 02 | May-2015.

[7] R. Ravikumar, Dr. V. Arulmozhi," Prediction of Leaf Diseases by using Machine Learning Techniques-A New Approach to Applied Informatics" International Journal on Recent and Innovation Trends in Computing and Communication Volume: 5 Issue: 6 , ISSN: 2321-8169 1154 – 1158.

[8] Milos Ilic, Petar Spalevic, Mladen Veinovic, Abdolkarim Abdala M. Ennaas, "Data mining model for early fruit diseases detection", In IEEE 23rd Telecommunications Forum Telfor (TELFOR), Jan. 2016

[9] Meenakshi, Geetika," Survey on Classification Methods using WEKA" International Journal of Computer Applications (0975 – 8887), Volume 86 – No 18, January 2014.

[10] Dataset: https://www.kaggle.com/vipoooool/new-plant-diseases-dataset